DONDERS
INSTITUTE

ISBN 978-94-6284-151-2

Max Planck Institute
for Psycholinguistics

Radboud University Radboudumc

EFFECTS OF COMPLEXITY IN PERCEPTION: FROM CONSTRUCTION TO RECONSTRUCTION

Yağmur Güçlütürk

DONDERS SERIES

330

Effects of complexity in perception: from construction to reconstruction

Yağmur Güçlütürk

Effects of complexity in perception: from construction to reconstruction

Yağmur Güçlütürk

ISBN
978-94-6284-151-2

Cover
Gustav Klimt

Effects of complexity in perception:
from construction to reconstruction

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus
prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op
vrijdag 29 juni 2018 om
10:00 uur precies door

Yağmur Güçlütürk

geboren op
8 januari 1987 te
Ankara (Turkije)

Promotor
Prof. dr. H. Bekkering

Copromotor
Dr. R.J. van Lier

Manuscriptcommissie
Prof. dr. R.J.A. van Wezel
Prof. dr. J. Wagemans (KU Leuven, België)
Prof. dr. C.C. Carbon (Otto-Friedrich-Universität Bamberg, Duitsland)

Effects of complexity in perception:
from construction to reconstruction

Doctoral thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus
prof. dr. J.H.J.M. van Krieken,
according to the decision of the Council of Deans
to be defended in public on
Friday, June 29, 2018 at
10:00 hours by

Yağmur Güçlütürk

Born on
January 8, 1987 in
Ankara (Turkey)

Supervisor
Prof. dr. H. Bekkering

Co-supervisor
Dr. R.J. van Lier

Doctoral Thesis Committee
Prof. dr. R.J.A. van Wezel
Prof. dr. J. Wagemans (KU Leuven, Belgium)
Prof. dr. C.C. Carbon (University of Bamberg, Germany)

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{99^2} \sum_{k=0}^{\infty} \frac{(4k)!}{k!^4} \frac{26390k + 1103}{396^{4k}}$$

— Ramanujan, 1914

Contents

1 Introduction **1**
1.1 Introduction 2
1.2 Outline 21

2 Liking versus Complexity: Decomposing the Inverted U-Curve **25**
2.1 Introduction 26
2.2 Materials and Methods 28
2.3 Results 33
2.4 Discussion 41

3 Decomposing Complexity Preferences for Music **53**
3.1 Introduction 54
3.2 Materials and Methods 56
3.3 Results 60
3.4 Discussion 68

4 Representations of Naturalistic Stimulus Complexity in Early and Associative Visual and Auditory Cortices **73**
4.1 Introduction 74
4.2 Materials and Methods 77
4.3 Results 90
4.4 Discussion 100

5 Convolutional Sketch Inversion **111**
5.1 Introduction 112
5.2 Materials and Methods 114
5.3 Results 120

5.4 Conclusion 128

6 Reconstructing perceived faces from brain activations with deep adversarial neural decoding 131
6.1 Introduction 132
6.2 Materials and Methods 133
6.3 Results 138
6.4 Conclusion 148

7 Summary and Discussion 151
7.1 Summary 152
7.2 Conclusion 164

Bibliography 167

Nederlandse samenvatting 201

Acknowledgements 203

Curriculum vitae 205

Donders Graduate School for Cognitive Neuroscience 211

1 Introduction

1.1 Introduction

Complexity is an essential concept with far-reaching implications in construction and reconstruction of percepts. Definition of complexity varies when viewed from different perspectives. While it is possible to obtain a formal definition of complexity from an information theory perspective, introducing perceptual processes into the equation makes it more difficult to achieve a consensus about what complexity means (Shmulevich & Povel, 2000).

In this section, various explanations of complexity in the contexts of information theory (Shannon, 1948), music perception and visual perception will be discussed, as the effects of complexity on perception of art and brain responses lies at the core of our investigations that are presented in the upcoming chapters.

Complexity from an Information Theory Perspective

Shannon's (1948) information theory was mainly developed with the purpose of achieving efficient transmission of messages (Chaitin, 1975). According to Shannon's information theory, unlikely and unpredictable patterns contain more information compared to predictable ones. Consider an example where the person at the receiver's end of a communication channel can predict the message even before he or she receives it. This would mean that he or she would receive no new information when the message arrives. Such a predictable message has low entropy and low information content according to information theory. An entirely random message which would be highly unpredictable, on the other hand, would have high entropy and have high information content. An important aspect of transmitting messages is to be able to represent the messages with the smallest possible length. This process of reducing a message (or any information) into a

smaller representation is called compression, and the complexity of a message can be defined as the compressibility of the original message.

We can illustrate the concept of compressibility with an example. Suppose that you would like to send the number π as a message to a friend. He said that he needs as many digits as possible after the decimal point because he is making a very sensitive and critical calculation. However, you can only communicate with your friend via SMS messages, and you do not want to spend all your time carefully entering numbers into your phone. If only there were a way to represent π in a shorter way! Then you think back to your geometry classes, and remember that π is in the formula of the circumference of a circle. Specifically, you remember that π is equal to the ratio of a circle's circumference to its diameter. However using this formula to calculate π would not be precise enough, since your friend might make some measurement errors in the process. Instead, you can send another formula by Ramanujan (1914), which is longer but allows calculating π with precision:

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{99^2} \sum_{k=0}^{\infty} \frac{(4k)!}{k!^4} \frac{26390k + 1103}{396^{4k}} \tag{1.1}$$

Although quite difficult to type, you decide to send this formula to your friend. He can then calculate π to the level of precision that he wants himself. Since you were able to reduce the infinite number of digits of π to merely a formula, we can conclude that π is highly compressible. However, if your friend asked for historical bitcoin price index rather than the number π, then you probably would not be able to come up with a such a compression, as it is almost random. And if you can observe some regularities which would allow you to compress it, then you could probably become rich!

Based on these concepts, in algorithmic information theory, Kolmogorov complexity of an object is defined as the length of the shortest computer program or algorithm that can be used to represent the object (Solomonoff, 1986). Estimations of Kolmogorov complexity has frequently been used to quantify complexity of images and similar digital materials using various file compression methods, such as ZIP, JPEG, GIF, etc. Using these methods the Kolmogorov complexity of an image is simply taken as its file size after compression (Donderi & McFadden, 2005).

Visual Complexity

The definition of visual complexity diverges from its definition in the context of information theory because, in the case of vision, perceptual limitations of humans (and other animals) and statistics of the environment that they have evolved in and live in have to be taken into account (Donderi, 2006; Sambrook & Whiten, 1997).

In the 20th century, Gestalt psychology contributed a great deal to the study of complexity (or rather simplicity) by looking for the relationship between sensory input and perceptual simplicity with the Gestalt laws of proximity, similarity, continuity, closure, and connectedness (Wertheimer, 1912; Köhler, 1920; Koffka, 1935). However, these studies failed to deliver an integrated account of these laws as to state the exact laws that apply to different patterns in a stimulus. In the pursuit of such an account, Koffka (1935) formulated the "Law of Simplicity" or "Law of Prägnanz", which postulated that experiences are ordered in a regular, orderly, symmetrical and simple manner. This law fell short on predicting perceived interpretations. Hochberg and McAlister (1953) offered a solution to this problem by combining Prägnanz with information to formulate the global minimum principle, which postulated that given a stimulus with multiple patterns, the one with the simplest possible interpretation would be perceived (Leeuwen-

berg, 1971; van der Helm & Leeuwenberg, 1991; van Lier, van der Helm & Leeuwenberg, 1994; van Lier, 1996). For two comprehensive reviews of Gestalt psychology in visual perception see (Wagemans et al., 2012a; Wagemans et al., 2012b). In the 1960s and 70s study of perceptual complexity has gained further traction largely owing to the experimental aesthetics studies which were led by Berlyne (Berlyne, 1963; Berlyne, Ogilvie & Parham, 1968; Berlyne, 1971) investigating the role of *collative variables*, among which complexity had a prominent place. According to Berlyne (1971), subjective complexity encompassed several different stimulus properties such as irregularity of arrangement, amount of material, heterogeneity of elements, irregularity of shape, number of independent units, asymmetry and random redistribution. Among these properties, number of elements and their heterogeneity were found to be the most crucial predictors of subjective complexity (Berlyne et al., 1968; Berlyne, 1971).

Neither Shannon's information theory nor algorithmic information theory takes the perceptually relevant regularities in the visual environment into account. Structural information theory (SIT) (Leeuwenberg, 1969; van der Helm, 2014; van Lier, 2001), on the other hand, was strongly influenced by the Gestalt psychology. Specifically, it was developed with the aim of explaining why some interpretations of visual stimuli were more preferred by the human visual system in the face of ambiguities. According to the Law of Simplicity of the Gestalt psychology, the visual system chooses the simplest explanation (or code) possible, and according to SIT, the simplest code is the code which can be formed using the least amount of information. This simplest code concept is analogous to the earlier illustrated idea of compression. However, in SIT a more restricted compression approach is adopted, such that the simplest code to reconstruct a visual stimulus needs to capture visual regularities in this stimulus in a hierarchical organization of wholes and parts. Although SIT captures the perceptual complexity of patterns and geometric shapes very well by

accounting for perceptually relevant regularities, it is less suitable for studying the complexity of natural images.

Subjective complexity of images have generally been measured by collecting complexity rating responses from a group of subjects for each image in a set of stimulus images, and these responses are often averaged across subjects to obtain an overall complexity estimate for each image. Since subjects have been shown to agree well on the complexity levels of images (Güçlütürk, Jacobs & van Lier, 2016b; Marin & Leder, 2013), averaging approach works well for measuring perceptual image complexity. However, this method can be costly, especially when the complexity levels of a large set of images are of interest, and can be easily confounded by subjective biases such as familiarity (Forsythe, Mulhern & Sawey, 2008). Alternatively, several different measures of image statistics, e.g. slope of the amplitude spectra (Spehar et al., 2015), fractal dimension (Taylor, Newell, Spehar & Clifford, 2005), self-similarity (Hayn-Leichsenring, Lehmann & Redies, 2017), histograms of oriented luminance gradients (HOG) (Hayn-Leichsenring et al., 2017), spatial envelope representation (GIST) (Oliva & Torralba, 2001), which are relevant in the fields of perception, psychophysics, and computer vision can be utilized to quantify the complexity of images. Besides being efficient and cheap compared to collecting subjective ratings, such measures also provide useful insights into properties of images, which are especially useful for natural images and artworks (Graham & Field, 2007; Geisler, 2008).

Another method of estimating perceptual complexity relies on the saliency maps of the images (Silva, Courboulay & Estraillier, 2011). This approach utilizes eye tracking data to calculate the saliency maps of images and quantify the amount of information in each map with Kolmogorov complexity. For such methods, computational visual saliency maps without the use of eye tracking data (Itti & Koch, 2001) can also be employed. Unlike other listed computational measures of complexity, saliency-based complexity

measures take the image content into account and attempt to mimic human attentional processes, which can be a beneficial measure, e.g. in human-computer interaction studies.

Returning to information theoretic perspective, Kolmogorov complexity of the algorithmic information theory still provides one of the easiest to obtain (estimated by the compressed file size of an image) and most accurate solutions to the computation/estimation of perceptual complexity. In the case of abstract images, as shown in Chapter 2 of this thesis, Kolmogorov complexity has been demonstrated to correlate almost perfectly with the perceptual complexity (i.e., the complexity of the visual stimuli as judged by our participants). Furthermore, it has been successfully used as a measure of performance in search tasks (Donderi & McFadden, 2005), and has been suggested as a reliable alternative to subjective complexity measures (Forsythe et al., 2008). Besides abstract images, it has been shown to correlate highly with subjective complexity of icons (Forsythe et al., 2008), artworks (Marin & Leder, 2016), photographs (Marin & Leder, 2013; Donderi & McFadden, 2005), websites (Tuch, Bargas-Avila, Opwis & Wilhelm, 2009) and computer displays (Donderi & McFadden, 2005). However as mentioned earlier, human perception of complexity is affected by factors not accounted for in this information theoretic perspective. For example, a random image, which has the least amount of compressibility and maximum complexity, may be perceived to be a rather simple stimulus (Grassberger, 1986; Sambrook & Whiten, 1997). According to Donderi (2006), our minds are not able to distinguish individual random patterns. Instead, we perceive each random pattern as the same image. This raises the overall probability of perceiving a random pattern (reducing its Shannon entropy, information content, and hence its complexity), making it perceptually simple.

Music Complexity

Considering that music has a temporal dimension, its complexity can be defined differently than visual complexity. This temporal dimension in combination with the importance of predictability and expectations in music makes it easier to relate music complexity to the information theoretic perspective (Meyer, 1957). As discussed earlier, according to Shannon's information theory, messages that have a high probability of occurrence, have less information content (and are less complex) compared to low probability messages (which are more complex). In music, this translates to songs that have highly predictable rhythm or melodies having a lower complexity level compared to those that have unpredictable rhythm or melodies.

However, despite these observed parallels, defining music complexity is difficult given its subjective nature (Shmulevich & Povel, 2000; Sallavanti, Szilagyi & Crawley, 2015). Besides the predictability of rhythm and melody, subjective music complexity has been determined to depend on several factors. That is, similar to its visual counterpart; it can be considered a multidimensional construct (Streich, 2007). Factors influencing the perceived complexity of music further includes harmony, frequency/number of events and variability instruments. While there are models of music complexity which take an integrated approach by combining multiple music elements (Mauch & Levy, 2011; Streich, 2007; Marin & Leder, 2013), several studies focused on defining the complexity of individual elements of music, such as rhythm and meter complexity (Shmulevich & Povel, 2000; Vuust & Witek, 2014; Thul & Toussaint, 2008), instrumental complexity (Percino, Klimek & Thurner, 2014), tonal complexity (Weiss & Muller, 2015), and harmonic complexity (Marsik, Pokorný & Ilcík, 2014).

While many music perception studies chose to vary the complexity of musical stimuli in selected dimensions and collect subjective rating measurements (Heyduk, 1975; Orr & Ohlsson, 2001; Orr,

2005; Madison & Schiölde, 2017), computational measures of music complexity have also been investigated. Examples of computational music complexity measures include entropy related measures, e.g., Shannon's entropy, entropy rate and excess entropy (Fleurian, Ben-Tal, Müllensiefen & Blackwell, 2014; Madsen & Widmer, 2006), Kolmogorov complexity, e.g., FLAC, Ogg Vorbis, MP3 file compression methods (Marin & Leder, 2013; Fleurian et al., 2014), event density (Marin & Leder, 2013), variability of temporal and spectral frequencies (Samson, Zeffiro, Toussaint & Belin, 2011b).

Role of Complexity in Perception and Cognition

Stimulus complexity is a stimulus property that has effects on performance in several different types of tasks. For example, visual stimulus complexity has been shown to influence time perception (Schiffman & Bobko, 1974), speed and accuracy of shape recognition (Kayaert & Wagemans, 2009), and the reaction time of search, discrimination and recognition tasks (Fitts, Weinstein, Rappaport, Anderson & Leonard, 1956; Anderson & Leonard, 1958; Mavrides & Brown, 1969; van Lier, 1998; Donderi & McFadden, 2005; Koning & Lier, 2005). Furthermore, studies showed its effects on memory (Simon, 1972; Chai, 2010), and attention and amodal completion in infants (Moffett, 1969; Richards, 2010; Vrins, Hunnius & van Lier, 2011). In Chapter 6 of this thesis, its effect on neural decoding of visual stimuli from brain activity is investigated. Complexity of auditory stimulus has also been shown to influence attention, arousal, and memory (Potter & Choi, 2006). Moreover, complexity of a song determines the context that it is used, e.g., complex songs are used for cognitive purposes, whereas simple ones are more likely to be used for emotional purposes (Sallavanti et al., 2015). Stimulus complexity also has a

strong influence on aesthetic feelings; these will be discussed in the following sections.

The processing of stimuli with varying levels of complexity may differ between individuals and throughout development. One such example is the differences between the sensory processing and sensitivities of people with and without autism spectrum disorder (ASD). For example, de Wit, Schlooz, Hulstijn and van Lier (2007) showed that children with ASD had differential priming effects in visual completion of complex occluded shapes. Furthermore, ASD has been linked to superior processing of fine details in the visual modality (Dakin & Frith, 2005), and observations suggest that they are more locally biased (rather than globally) during visual processing (Vanmarcke, Noens, Steyaert & Wagemans, 2017). Similarly, in the auditory domain, individuals with ASD have been observed to possess normal or enhanced abilities of identification and discrimination of isolated low-level perceptual acoustic features, which may have consequences on perception of complex speech stimuli (Haesen, Boets & Wagemans, 2011). Atypical processing of complex sensory stimuli in ASD also extends to multisensory integration (Kwakye, Foss-Feig, Cascio, Stone & Wallace, 2011).

Other factors influencing complexity perception include age (Chai, 2010), psychiatric health (Iakovides et al., 2004), language skills (Roncaglia-Denissen, Roor, Chen & Sadakata, 2016) and expertise (Vogt & Magnussen, 2007; Koide, Kubo, Nishida, Shibata & Ikeda, 2015; Schlaug, 2006; Besson, Schön, Moreno, Santos & Magne, 2007; Chamberlain & Wagemans, 2015). For example, as a result of training, musicians may obtain superior pitch perception (in both music and language) compared to non-musicians, which is also reflected in differences in structural and functional differences in brain regions that are known to process music and language (Schlaug, 2006; Besson et al., 2007). In the visual domain, art expertise relates to differences in attention to local and global stimulus features (Chamberlain & Wagemans, 2015), as well as differ-

ences between eye-movement patterns, when viewing both real-world scenes and artworks (Vogt & Magnussen, 2007; Koide et al., 2015).

Complexity in Art

Art styles differ from each other in terms of several characteristics including their complexity levels. For example, considering different artistic styles of paintings, one would find differences in perceptual attributes, such as color saturation and temperature, depth, complexity, types of brush strokes as well as differences in more content related attributes, such as the level of abstraction, realism, animacy, realism, emotional expressivity and symbolism (Chatterjee, Widick, Sternschein, Smith & Bromberger, 2010). Furthermore, artworks can also be classified in terms of the used artistic medium, which may have effects on the visual complexity of the artwork. For example, a pencil sketch is by definition a simpler version of an oil painting or a photograph, such that it contains less information than the painting, particularly in terms of the color and detail content.

Concerning different artistic styles that vary in terms of their complexity levels, we investigated transformations between simple and complex art types in Chapter 5 of this thesis. Specifically, we first converted photographs to face sketches using a simple computational filtering method. Then, we converted the sketches back to photographs using deep neural networks, which are able to learn relations between input-output pairs when shown enough examples. The reconstructions almost perfectly matched the actual photographs. In the context of information theory, such an application presents an example of a (lossy) compression algorithm, such that the grayscale and low-size sketch is the compressed representation of the original color image, given that the sketch could be used to recreate the original image. This is similar to

what artists do when they combine sketching with digital coloring, i.e. first sketching the scene on the spot in as much detail as possible, and later digitally coloring the sketches relying on their memory and general knowledge.

In the case of the classification of the music genres, complexity can again be considered as an important factor. For example, based on the "axiomatic triangle" of Philip Tagg, music can be classified into three genres: art, pop, and folk music (Tagg, 1982). According to this classification, a main difference between the art and pop music genres is that pop music is accessible and simple whereas art music demands for constant focused attention and is complex. However, this categorization offers a rather unspecified division of the vast scope of music styles. From a psychological perspective, to explain music in terms of preferences, Rentfrow, Goldberg and Levitin (2011) developed the MUSIC theory and identified five main dimensions of music (Mellow, Unpretentious, Sophisticated, Intense, Contemporary) that comprise several psychological and sonic attributes spanning the whole music spectrum. Music genres can further be thought of in terms of their instrumental (Percino et al., 2014) or computationally measured complexity levels (Cataltepe, Yaslan & Sonmez, 2007).

It is also possible to characterize different artworks and artistic styles in terms of their statistical image properties that relate to the overall complexity of images. Such measures include amplitude spectra, fractal dimension, self-similarity, among others. Amplitude spectra is a measure of the distribution of spatial frequencies in the image, which is obtained by applying Fourier analysis. While natural images and western artworks are known to share similar characteristics in terms of their amplitude spectra, face photographs and drawings differ in their average amplitude spectra slopes (Graham & Field, 2007; Koch, Denzler & Redies, 2010). Fractal dimension and self-similarity are two complementary computational measures, which are both related to patterns and irregularities in images at different scales. Specifically, fractal

dimension of an image quantifies how much self-similarity it contains across different spatial scales, where self-similarity is a measure of how much the whole of an object resembles its parts (Mandelbrot, 1977). Carballal et al. (2016) showed that fractal dimension, among other complexity related image features such as compressed file size, could be used to train an artificial neural network to determine whether an image was a painting or a photograph. Fractal dimension has been further used as a predictor of beauty of different art styles (Forsythe, Nadal, Sheehy, Cela-Conde & Sawey, 2011; Spehar, Clifford, Newell & Taylor, 2003). Pollock's drip paintings are a famous example of fractal dimension in art. His paintings have been shown to carry changing fractal characteristics as he produced paintings with consistently increased fractal dimensions over the years (Taylor, Micolich & Jonas, 1999). Next to fractal dimension, several recent studies (Redies, Amirshahi, Koch & Denzler, 2012; Braun, Amirshahi, Denzler & Redies, 2013; Hayn-Leichsenring et al., 2017) analyzed and compared artworks with measures derived from histograms of oriented luminance gradients (Bosch, Zisserman & Munoz, 2007). They found such statistical image properties to be useful for categorical characterization as well as being correlated with liking and aesthetic evaluations of paintings (Hayn-Leichsenring et al., 2017).

Complexity relates to interestingness and pleasantness as well as other aesthetic evaluations of art (Berlyne, 1963; Berlyne et al., 1968). Similarly from a structural information perspective, beauty relates to the surplus of regularity, i.e. the regularity not necessary describe the visual pattern (Boselie & Leeuwenberg, 1985). Complexity has been established as an important factor in art perception and has a critical role in current models of art perception and aesthetic appreciation (Leder, Belke, Oeberst & Augustin, 2004; Leder & Nadal, 2014; Chatterjee, 2004; Redies, 2015; Muth, Raab & Carbon, 2015, 2016). Furthermore, as it is the case with visual arts, music complexity profoundly influences whether a song will be liked or not (Orr, 2005; Marin & Leder, 2013, 2016;

Heyduk, 1975; North & Hargreaves, 1995). Berlyne (1971) suggested that an inverted U-curve could explain the relationship between the collative variable complexity and how pleasant a stimulus feels. This inverted U-curve relationship would entail that the stimuli with intermediate levels of complexity would be the most preferable ones. In support of this theory, many empirical aesthetics studies – both in visual arts and music – found an inverted U-curve when characterizing aesthetic feelings as a function of complexity (Vitz, 1966; Saklofske, 1975; Heyduk, 1975; Farley & Weinstock, 1980; North & Hargreaves, 1995; Imamoglu, 2000). Complexity has since been much studied in the domain of art perception and empirical aesthetics for its role in preferences, with many studies finding variations of complexity and liking relationship, which do not conform to the inverted U-curve function (Nadal, Munar, Marty & Cela-Conde, 2010; Güçlütürk et al., 2016b; Marin & Leder, 2013).

Individual Differences in Complexity Preferences

Liking and aesthetic emotions are predominantly subjective and dynamic (Carbon, 2010, 2011). This means that preferences of individuals not only differ from others' preferences but also even from their own choices, moment to moment. Therefore it is not surprising to see that variability and large individual differences have always been the part of empirical studies of liking and aesthetic preferences (Nusbaum & Silvia, 2010; Palmer & Griscom, 2012). As early as in 1917, Thorndike called attention to the subjectivity of preferences by highlighting the fact that the average tendency of a group does not necessarily imply the existence of a uniform agreement among the individuals that make up this group (Thorndike, 1917). He studied the preferences for shapes of rectangles (among other simple geometric shapes),

and found that while the average tendency of his sample was towards a preference for height to base ratios close to the golden ratio in rectangles, there were also large variations among the preferences of individuals for simple geometric shapes. In the case of complexity preferences, large differences between participants were reported in several studies (Vitz, 1966; Crosson & Robertson-Tchabo, 1983; Aks & Sprott, 1996; Jacobsen & Höfel, 2002; McManus, Cook & Hunt, 2010; Nadal et al., 2010; Spehar et al., 2015). Very similar to what was reported by Thorndike (1917), McManus et al. (2010) found differences in preferences for rectangle side ratios. Similarly, in another study investigating complexity and symmetry preferences, it was found that while a group-level analysis pointed towards a general preference for high complexity patterns, individual participants with the opposite preference, who liked simple patterns more, were also present among the same set of participants (Jacobsen & Höfel, 2002).

A recent contribution to the study of complexity preferences has been provided by investigations of discrepant results in the field (Nadal et al., 2010). Their study showed that complexity-liking relationship could vary to show an inverted U-curve, a (non-inverted) U-curve, or linear increase based on the type of complexity dimension of the stimuli that were systematically manipulated. For example, they found that as the number of elements in an artwork increased, the beauty ratings also increased in an almost linear relation. The explanation by Nadal et al. (2010) relates specifically to the various uses of the term complexity among different studies. However, the vagueness of the adopted terminology in the field of empirical aesthetics, in general (Augustin, Wagemans & Carbon, 2012a; Augustin, Carbon & Wagemans, 2012b), is also highlighted by such results.

Besides the identification of groups of people with different preferences, a principal aim of empirical aesthetics should be to identify the factors related to the heterogeneity of preferences. Such factors may be among more static or slowly changing variables

such as age, sex, education, occupation, personality, sensory sensitivity, as well as among more dynamic variables such as understanding of the stimuli, mood and motivation. Age differences have been previously identified to contribute to complexity preference differences, such that older age was related to a tendency toward a preference for simpler visual stimuli (Munsinger, Kessen & Kessen, 1964; Alpaugh & Birren, 1977; Crosson & Robertson-Tchabo, 1983). Another factor that relates to complexity preferences is frequency of visits to galleries or museums, which is positively correlated to preference for complexity (Chamorro-Premuzic, Burke, Hsu & Swami, 2010; Cleridou & Furnham, 2014). A recent promising direction in understanding individual differences in preferences is the investigation of a biological basis of such differences. For instance, Spehar et al. (2015) showed large correspondences between visual sensitivity and liking of spectral noise and sinusoidal grating patterns. A good example of the influence of dynamic variables in the variability of preferences and aesthetic feelings is the 'Aesthetic Aha' moment, in which an insight regarding an ambiguous artwork or image is achieved (Van de Cruys & Wagemans, 2011; Muth & Carbon, 2013). Such moments may be interpreted as achieving a simpler interpretation of the image, i.e. Gestalt detection, and often result in a sudden increase in liking of the viewed artwork (Muth & Carbon, 2013).

In the topic of music appreciation, in a recent study, Marin and Leder (2013) investigated sex differences in complexity preferences for music and visual stimuli. They observed a larger preference for complexity among males compared to females for piano solo and piano trio excerpts. In fact, while the rated pleasantness and complexity of excerpts were negatively correlated for female participants, it was positively correlated for the male participants. As mentioned earlier, music styles and genres have been well studied in terms of preference differences (Rentfrow et al., 2011; Rentfrow et al., 2012; Greenberg, Baron-Cohen, Stillwell, Kosinski & Rentfrow, 2015). In these studies, personality, age and sex have all been identified as important factors. Concerning

instrumental complexity of music, it has been shown that the number of the album sales were negatively correlated with the instrumental complexity of the songs in an album (Percino et al., 2014). This indicates a larger tendency in the population towards instrumentally less diverse music.

In Chapters 2 and 3 of this thesis, we look into individual differences in complexity preferences. In Chapter 2, using clustering analysis, we investigate preferences for visual stimuli of algorithmically generated digital art, which has fractal-like properties (Shier, 2011). In Chapter 3, we extend our investigations to the auditory domain and replicate our findings in the visual domain for ecologically valid music stimuli spanning a large part of the musical spectrum.

Complexity Representations in the Brain

Based on early animal (Hubel & Wiesel, 1962; Gross, Rocha-Miranda & Bender, 1972; Tanaka, 1996) and human fMRI studies (Güçlü & van Gerven, 2015; Güçlü, Thielen, Hanke & van Gerven, 2016), we know that brain regions in the sensory cortices code for increasingly more complex stimulus features along the sensory hierarchy. However, despite its importance, studies investigating the perception of complexity of naturalistic stimuli in the brain have been limited. In the earliest days of functional brain imaging using positron emission tomography (PET), researchers investigated differences between the brain responses to diffuse light as the simplest stimulation option, and checkerboard pattern and a movie as complex stimulation patterns (Phelps, Kuhl & Mazziota, 1981; Sitzer, Diehl & Hennerici, 1992). The study by Phelps et al. (1981) showed that both simple and complex visual stimuli caused changes in the cerebral metabolic rate for

glucose in the primary visual cortex and associative visual cortices. However, they observed that the changes caused by the more complex stimulation, i.e. the checkerboard pattern, had resulted in a faster change of cerebral metabolic rate for glucose in the associative visual cortex. Sitzer et al. (1992) found that increasing the complexity of visual stimulation caused the velocity of blood flow in the occipital cortex to increase. Another relevant study investigating the effects of visual stimulus complexity in the human brain showed correspondences between the amplitude of N2 event related potential and complexity of visual stimulus (Shigeto, Ishiguro & Nittono, 2011). The complexity of their visual stimulus was defined as the number of edges that randomly generated polygons had. In a multimodal study investigating the influence of complexity on emotional stimulus processing using fMRI, the researcher defined the levels of complexity as words, phrases, pictograms, and photographs (Schlochtermeier et al., 2013). They measured the brain activity of participants during a valence judgment task using emotional stimulus with different complexity levels, and besides modality-specific activity differences, they found no differences in terms of emotion processing in the brain.

Samson et al. (2011b) studied complexity processing in the auditory cortex with a meta-analysis of studies that reported neural activity in response to several categories of auditory stimulus. They defined auditory stimulus complexity in terms of different categories of sound. From simpler to complex these categories were as follows: pure tones, noise, music and vocal sounds. They presented an overview of where in the human auditory cortex most activity in response to these different types of auditory stimuli has been reported. In a separate study, Samson et al. (2011a) investigated differences in processing of temporal and spectral auditory complexity between individuals with ASD and controls. Temporal and spectral complexity in this study were defined as the amount of spectral and temporal variability, respectively. They found no differences in the processing of spectral complexity, but

greater response to temporal complexity in the primary auditory cortex in ASD and reduced activity in the remaining regions of the auditory cortex. More recently, Wilkins, Hodges, Laurienti, Steen and Burdette (2014) investigated brain connectivity patterns of participants who were listening to their favorite songs. They found that even when the complexity levels of the songs that participants enjoyed were greatly different, the connectivity patterns were similar.

In Chapter 4 of this thesis, we investigated the neural representations of complexity in the brain with fMRI. Particularly, in two separate experiments, we looked at the complexity representations in the visual and auditory cortices of participants while they viewed or listened to naturalistic stimuli (photographs or music excerpts).

Decoding Complex Stimuli from the Brain

Another area of study that stimulus complexity plays an important role is decoding brain activity, i.e. reconstructing perceived stimuli from measured brain activity. Initial attempts of decoding perceived stimuli from the human brain started with simple decoding tasks such as identifying the perceived stimulus category such as faces, cats, man-made objects, and non-sense pictures (Haxby, 2001). Following the success in classification of categories of percepts, later studies investigated methods to reconstruct the perceived images of simple stimuli such as geometric shapes (Thirion et al., 2006), letters and numbers (Miyawaki et al., 2008; van Gerven, de Lange & Heskes, 2010; Schoenmakers, Barth, Heskes & van Gerven, 2013). These studies mostly relied on retinotopy and the activity patterns of voxels in the early visual cortex, and their reconstructions matched the stimuli very well. More complex stim-

uli such as natural images were reconstructed to a lesser degree of accuracy in terms of the similarity of reconstructions to the target stimuli (Naselaris, Prenger, Kay, Oliver & Gallant, 2009). Compared to simpler stimuli, such as geometric shapes or letters, the signal to noise ratio of stimulus-evoked voxel responses to natural images is lower. This is a factor that makes the reconstruction of natural images a more difficult problem compared to simpler stimuli (Naselaris et al., 2009). The method of Naselaris et al. (2009) relied on first predicting the voxel responses to an extensive set of images using an encoding model, and then selecting the stimuli whose predicted response pattern was most similar to the stimulus-evoked voxel responses of the target stimulus. Extending the scope of the studies from static images to dynamic stimuli, Nishimoto et al. (2011) then showed that it was possible to reconstruct the gist of movies from stimulus-evoked voxel responses. They also studied motion processing in the early visual areas and showed dependencies between speed tuning and eccentricity. Similar to Naselaris et al. (2009), Nishimoto et al. (2011) also utilized an encoding model to predict voxel responses to a very large set of movies (which were not part of the fMRI experiment) and then selected the most similar ones to the actual stimulus-evoked responses and averaged these selected movies to obtain reconstructions. As expected, this method results in very blurry reconstructions due to the averaging of movies.

Faces are an essential type of stimuli that are socially valuable for primates, which is evident by the presence of brain regions specialized to process faces both in human and non-human primates (Yovel & Freiwald, 2013). Previous decoding approaches that were reviewed above, which mostly relied on retinotopy and V1 activity, would neither leverage the specialized and high level representations of faces in the human brain nor take advantage of the structured nature of the faces. For these purposes, Cowen, Chun and Kuhl (2014) used principal component analysis (PCA) to first identify representative face features. Using partial least squares regression, they mapped these identified face features to

patterns of fMRI activity, and used this mapping to reconstruct faces from stimulus-evoked voxel responses. For reconstructions, they used voxels from different brain regions, such as the occipital cortex and fusiform face area (FFA), and found that reconstruction accuracy was the highest when only the voxels in the occipital cortex were used. Combining the voxels in FFA and occipital cortex also produced good results. While their reconstructions were both subjectively and objectively identifiable among a set of face images above chance level, they lacked realistic sharp features. Recently, using a similar approach to Cowen et al. (2014), Lee and Kuhl (2016) showed that voxel responses in angular gyrus of the lateral parietal cortex had representations of perceived and remembered faces.

In Chapter 6, we present state-of-the-art face reconstruction results from human brain activity using deep neural networks. Our reconstructions look realistic and match the target faces in terms of essential characteristics such as gender, skin color, and facial structure. After accomplishing this high performance decoding method, we investigate the role of complexity in reconstruction performance, and reveal it as an important factor.

1.2 Outline

The remainder of this thesis is organized into two parts. In the first part, we present three different studies dealing with the effects of complexity in the construction of percepts, in Chapters 2, 3 and 4. In the second part, we present two studies demonstrating reconstructions of complex visual stimuli from simpler images and stimulus-evoked brain responses, in Chapters 5 and 6, respectively. Below, we briefly outline the studies that will be presented in each chapter.

The relationship between liking and stimulus complexity is commonly reported to follow an inverted U-curve. However, large individual differences among complexity preferences of participants have frequently been observed since the earliest studies on the topic. The common use of across-participant analysis methods that ignore these large individual differences in aesthetic preferences gives an impression of high agreement between individuals. In Chapter 2, we collected ratings of liking and perceived complexity from 30 participants for a set of digitally generated grayscale images. In addition, we calculated an objective measure of complexity for each image. Our results reveal that the inverted U-curve relationship between liking and stimulus complexity comes about as the combination of different individual liking functions. Specifically, after automatically clustering the participants based on their liking ratings, we determined that one group of participants in our sample had increasingly lower liking ratings for increasingly more complex stimuli, while the second group of participants had increasingly higher liking ratings for increasingly more complex stimuli. Based on our findings, we call for a focus on the individual differences in aesthetic preferences, adoption of alternative analysis methods that would account for these differences and a re-evaluation of established rules of human aesthetic preferences.

The previous chapter demonstrates complexity as a major factor in explaining individual differences in visual preferences for abstract digital art. In Chapter 3, building upon these results, we extend our investigations for complexity preferences from highly controlled visual stimuli to ecologically valid stimuli in the auditory modality. Similar to visual preferences, we show that music preferences are highly influenced by stimulus complexity. We demonstrate this by clustering a large number of participants based on their liking ratings for song excerpts from various musical genres. Our results show that, based on their liking ratings, participants can best be separated into two groups: one group with a preference for more complex songs and another group

with a preference for simpler songs. Finally, we consider various demographic and personal characteristics to explore differences between the groups, and report age and gender as significant factors separating the two groups.

The complexity of sensory stimuli has an important role in perception and cognition. However, its neural representation is not well understood. In Chapter 4, we characterize the representations of naturalistic visual and auditory stimulus complexity in early and associative visual and auditory cortices. This is realized by means of encoding and decoding analyses of two fMRI datasets in the visual and auditory modalities. Our results implicate most early and some associative sensory areas in representing the complexity of naturalistic sensory stimuli. For example, PPA, which was previously shown to represent scene features, is shown to also represent scene complexity. Similarly, posterior STG/S, which was previously shown to represent syntactic (language) complexity, is shown to also represent music (auditory) complexity. Furthermore, our results suggest the existence of gradients in sensitivity to naturalistic sensory stimulus complexity in these areas.

In Chapter 5, we use deep neural networks for inverting face sketches to synthesize photorealistic face images. We first construct a semi-simulated dataset containing a very large number of computer-generated face sketches with different styles and corresponding face images by expanding existing unconstrained face data sets. We then train models achieving state-of-the-art results on both computer-generated sketches and hand-drawn sketches by leveraging recent advances in deep learning such as batch normalization, deep residual learning, perceptual losses and stochastic optimization in combination with our new dataset. We finally demonstrate potential applications of our models in fine arts and forensic arts. In contrast to existing patch-based approaches, our deep-neural-network-based approach can be used for synthesizing photorealistic face images by inverting face sketches in the wild.

In Chapter 6, we present a novel approach to solve the problem of reconstructing perceived stimuli from brain responses by combining probabilistic inference with deep learning. Our approach first inverts the linear transformation from latent features to brain responses with maximum a posteriori estimation and then inverts the nonlinear transformation from perceived stimuli to latent features with adversarial training of convolutional neural networks. We test our approach with a functional magnetic resonance imaging experiment and show that it can generate state-of-the-art reconstructions of perceived faces from brain activations.

2

Liking versus Complexity: Decomposing the Inverted U-Curve

This chapter is based on Güçlütürk, Y., Jacobs, R., and van Lier, R. (2016). Liking versus complexity: decomposing the inverted U-curve. Frontiers in Human Neuroscience, 10:112. https://doi.org/10.3389/fnhum.2016.00112

2.1 Introduction

The relationship between the complexity of a stimulus and its perceived beauty has been a topic of great interest with influential studies since the earlier experimental investigations of aesthetics. For instance, Berlyne showed that complexity is a dominant determinant of interestingness and pleasingness of a stimulus (Berlyne, 1963; Berlyne et al., 1968). Berlyne (1971) suggested that the relationship between complexity and pleasingness could be explained by an inverted U-curve, where the stimuli with intermediate levels of complexity are the most preferable ones. This concept of an optimal amount of stimulus complexity has been supported by numerous studies that found an inverted U-curve when characterizing aesthetic preference as a function of complexity (Vitz, 1966; Berlyne, 1971; Saklofske, 1975; Farley & Weinstock, 1980; Imamoglu, 2000).

Although, Berlyne's theory has been influential in the field of experimental aesthetics (Silvia, 2005), Nadal et al. (2010) point out some discrepancies among the results of previous studies investigating the relationship between liking and complexity. Nadal et al. (2010) illustrate that studies utilizing a systematic manipulation of the degree of stimulus complexity resulted not always in an inverted U-shaped characterization of aesthetic preference as a function of complexity, but sometimes increasing, decreasing or U-shaped characterizations of the relationship. Nadal et al. (2010) suggested that the results vary since different studies manipulated different complexity dimensions, and different complexity dimensions have different relationships with aesthetic preference. Although, we agree that the various complexity dimensions utilized in these studies could have a major effect on the divergence of the results, an additional factor contributing to this divergence could be the individual differences in aesthetic preferences of the participants. Large individual differences among complexity preferences of participants have been frequently observed even on a

single study level (Vitz, 1966; Crosson & Robertson-Tchabo, 1983; Aks & Sprott, 1996; Jacobsen & Höfel, 2002). For instance, Jacobsen and Höfel (2002) found that, while the group-level analysis indicated a preference for higher levels of complexity, their sample of participants also included individuals who displayed exactly the opposite complexity preference patterns. These examples make it clear that it is desirable to look more carefully at individual differences. Individual differences in the experimental aesthetics literature are of course not limited to preference differences in complexity. More recently, a number of studies explored the relationship between characteristics of individuals (e.g., age, gender, educational background, and personality traits) and various types of aesthetic preferences such as preference for paintings of various artistic styles (Chamorro-Premuzic, Reimers, Hsu & Ahmetoglu, 2009; Chamorro-Premuzic et al., 2010; Cleridou & Furnham, 2014), preference for architecture and music of various artistic styles (Cleridou & Furnham, 2014), preference for rectangles with various side length ratios (McManus et al., 2010), strength of musical aesthetic experiences (Nusbaum & Silvia, 2010) and harmony preference (Palmer & Griscom, 2012). The results of these studies were not entirely consistent with each other, such that for example, while Chamorro-Premuzic et al. (2009), Chamorro-Premuzic et al. (2010), and Cleridou and Furnham (2014) found correlations between various personality traits and aesthetic preferences, McManus et al. (2010), and Palmer and Griscom (2012) did not. Such inconsistencies between the results of aesthetic studies further emphasize the need for systematically investigating these preference differences.

Besides studying the preference differences on an individual level, we suggest that identifying subgroups of participants with similar preferences and then characterizing these subgroups can result in much progress in experimental aesthetics. To embark on such a task we suggest taking an exploratory approach and using clustering analysis. Since their initial use in psychology by Zubin (1938), clustering analysis methods have been embraced as a means of

connecting the study of individuals (idiographic approach) and the study of cohorts of individuals (nomothetic approach), for example in health psychology, a field in which gaps between theory and individual cases are common (Clatworthy, Buick, Hankins, Weinman & Horne, 2005). Similarly, the differences between the nomothetic and idiographic approaches have been a concern in the study of aesthetics (Berlyne, 1977; Jacobsen & Höfel, 2002; Jacobsen, 2004; Mallon, Redies & Hayn-Leichsenring, 2014). Utilization of a clustering approach in experimental aesthetics would allow detecting subgroups of individuals having different patterns of preferences in relation to variable(s) of interest, e.g., complexity, symmetry, color, etc. After detecting these subgroups, shared characteristics (e.g., personality traits, mood, cultural background, art education, etc.) of the individuals belonging to particular subgroups can be further investigated to increase our understanding of the mechanisms underlying the aesthetic preferences.

In this study, adopting an exploratory and assumption-free approach rather than a hypothesis driven one, we aimed to demonstrate the existence of differences in complexity preferences of people, and argue against a universal rule of inverted U-curve for explaining the complexity-liking relationship. Concretely, using abstract computer-generated art as stimuli, we first show that liking versus complexity ratings averaged across all participants result in an inverted U-curve as often found in the literature. Next, by clustering participants based on their liking rating patterns, we demonstrate that this inverted U-curve is a result of combining two different liking versus complexity functions from two groups of participants. Finally, we show response time differences between these two groups of participants and discuss the implications of our results.

2.2 Materials and Methods

Participants

Thirty participants (average age ± standard deviation: 21.3 ± 3.5, 19 female) participated in the experiment in exchange for course credit or monetary compensation. They all had normal or corrected to normal vision. The study was approved by the Radboud University Ethics Committee Faculty of Social Sciences, and all participants provided written informed consent in accordance with the Declaration of Helsinki.

Stimuli

As stimuli, we generated statistical geometric patterns (SGPs) using a space filling algorithm that randomly places non-overlapping geometric shapes that monotonically decrease in size (Shier, 2011; Shier & Bourke, 2013). The algorithm used in the present study was originally developed for generating art using computer programs by Shier (2011). For various examples of his 'algorithmic art,' the reader is referred to Shier's website at http://john-art.com.

The stimulus set consisted of 144 grayscale SGP images of basic geometric shapes; namely circle, hexagon, square, and triangle (see Figure 2.1*A* for a subset of the stimulus set, and supplementary materials for the complete set of stimuli used in the experiment). Each SGP image was built up by filling a square surface with same-shaped elements. The value of a parameter called 'c' determined both the size of the first shape element and the speed with which this size decreased (note that the parameter 'c' is a variable used in the generation of the stimuli, but not a measure of complexity *per se*. Further on we will relate 'c' to perceptual complexity). The square was filled up with element shapes until 55% of its surface area was filled, or until 5000 ele-

Stimulus material and data are available upon email request from the corresponding author.

ment shapes had been placed. The elements did not overlap. The filled-up surface had a mid-level gray color, and each element had a random gray level. We generated 36 stimuli per geometric shape with equally spaced *c*-values ranging from 0.2 to 1.7. Stimuli were saved as and presented in Portable Network Graphics (PNGs) file format.

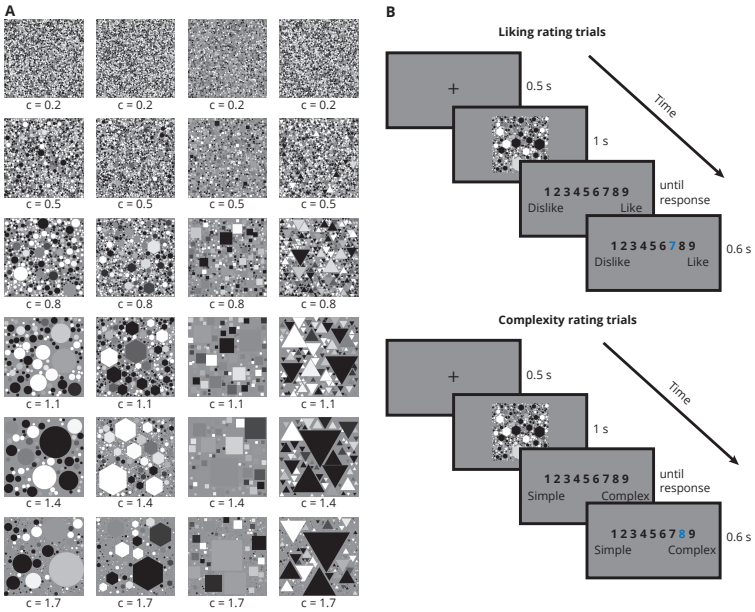

Figure 2.1: Example stimuli and experiment design. ***A***: Example stimuli with shape elements circle, hexagon, square, and triangle. ***B***: Timeline of the liking and complexity rating trials.

Procedure and Design

The experiment started with a short training session where participants gave nine-point Likert scale liking ratings for images presented on a computer screen. Participants were informed that the lowest rating on the scale (one) meant 'I do not like it at all' and the highest rating on the scale (nine) meant 'I like it a lot,' whereas a rating of five was neutral, and similarly for complexity ratings one meant 'very simple' and nine meant 'very complex.'

No further definitions of liking or complexity were provided. Participants were encouraged to try and use the whole scale for their ratings. Through the training session, participants became familiar with the task and the type of stimulus to expect during the rest of the experiment. The stimuli presented during the training session were not used during the main rating sessions. The rest of the experiment consisted of two rating sessions: liking and complexity (Figure 2.1*B*). At the beginning of each session, an instruction screen reminded participants either to give ratings for how much they like the images or how complex they find the images. Every trial started with a fixation cross in the middle of a mid-level gray screen, which remained there for 500 ms, followed by the stimulus presentation for 1 s. This presentation duration is in line with the recent work by Marin and Leder (2016), who showed that presentation durations of 1 and 5 s result in very similar complexity and pleasantness ratings. After the stimulus presentation a rating screen came up, displaying a scale of numbers from one to nine. The rating screen stayed on until the response by the participant. Participant responses were entered through number keys on a computer keyboard. Once the evaluation was made by the participant, the selected number was highlighted for 600 ms before the next trial started. Each rating session consisted of 144 trials in which participants gave ratings to all stimuli. Each stimulus was presented only once per session and the stimulus presentation order was randomized for each participant. In order to avoid possible differential effects of familiarity on the liking results, liking sessions always took place before the complexity sessions.

Analysis

Clustering of Participants Based on Their Liking Ratings

In this study, the k-means clustering algorithm (Lloyd, 1982) was used to cluster the participants into several groups (ranging from 2 to 8) based on their liking ratings averaged across stimuli having the same *c*-value. Cluster analysis groups individual elements in a set in such a way that the similarity of elements assigned to the same group is maximized, whereas the similarity between different groups is minimized. Most often, similarity is defined in terms of a distance measure between the elements. K-means clustering algorithm is the simplest and most used clustering algorithm (Rokach, 2009). In this algorithm each cluster is represented by its center, which is calculated as the mean of all data points in that cluster. Starting with an initial set of cluster centers, the algorithm iteratively partitions data into k clusters by assigning each data point to a cluster such that the within cluster sum of squares is minimized. This is equivalent to assigning each data point to the nearest cluster.

We implemented the k-means clustering algorithm to cluster participants based on their liking ratings using the *kmeans* function of MATLAB with squared Euclidean distance measure. Cluster centers were initialized with the *k-means++* algorithm (Arthur & Vassilvitskii, 2007) as implemented in MATLAB.

In order to determine the optimum number of clusters, the average silhouette values across all data were calculated for each value of k, i.e., for $k = 2, 3, 4, 5, 6, 7,$ and 8. The silhouette value measures a data point's within cluster similarity in comparison to its between cluster similarity (Rousseeuw, 1987). A large average silhouette value implies a better clustering of the data, i.e., it implies that the distances between the participants in the same cluster were minimized while the separation between different

clusters was maximized. Silhouette values were calculated with squared Euclidean distance measure using the *silhouette* function of MATLAB.

2.3 Results

First, the liking ratings were normalized, i.e., the z-score of each rating per participant was calculated. Z-scoring brings the ratings of the individual participants to the same scale, while preserving the relative distances between individual ratings and the shape of the data (Glenberg, 2008; Mitsa, 2010). This scaling in turn allows a comparison of normalized ratings by eliminating possible confounds in the data that stem from the rating style of the participants. An example of confounds that z-scoring helps eliminate is the differences between the participants in terms of their use of the range of the rating scale (some confine their ratings to the middle of the range whereas some use the entire range). Another example is the response biases (a general tendency to give high or low ratings) of the participants. Figure 2.2*A* shows the normalized liking ratings for each circle, hexagon, square, and triangle SGP stimuli averaged across all participants versus the *c*-values of the stimuli. The liking ratings and the *c*-values were significantly correlated for all shapes (Pearson product-moment correlation coefficient $r_{\text{circle}} = 0.481$, $p_{\text{circle}} = 0.003$; $r_{\text{hexagon}} = 0.839$, $p_{\text{hexagon}} < 0.001$; $r_{\text{square}} = 0.481$, $p_{\text{square}} = 0.003$; $r_{\text{triangle}} = 0.424$, $p_{\text{triangle}} = 0.010$). Similarly, Figure 2.2*B* shows the normalized complexity ratings for each circle, hexagon, square, and triangle SGP stimuli averaged across all participants versus the *c*-values of the stimuli. Overall, as the *c*-value increased, the complexity ratings decreased (Pearson product-moment correlation coefficient $r_{\text{circle}} = -0.810$, $p_{\text{circle}} < 0.001$; $r_{\text{hexagon}} = -0.900$, $p_{\text{hexagon}} < 0.001$; $r_{\text{square}} = -0.815$, $p_{\text{square}} < 0.001$; $r_{\text{triangle}} = -0.850$, $p_{\text{triangle}} < 0.001$), however, a non-linear relationship between the complexity ratings and the *c*-values was visible for all shapes. In the further analysis, the

main focus will be on measures of complexity (subjective and objective) and liking, but not the *c*-values.

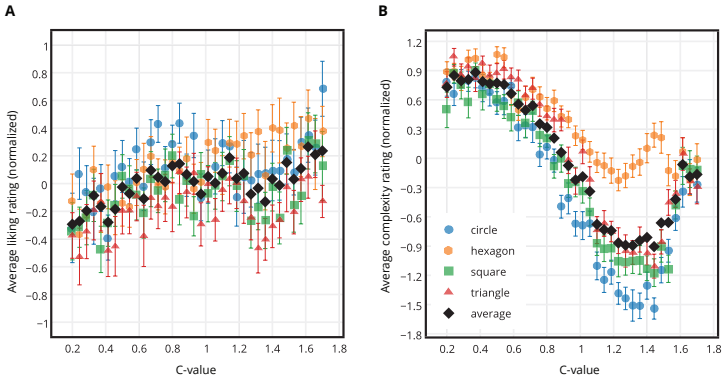

Figure 2.2: Liking and complexity ratings for each shape type versus the *c*-values. ***A***: Normalized liking ratings for circle, hexagon, square, and triangle SGP stimuli averaged across participants (colored dots), and across participants and shapes (black dots) versus the *c*-values. ***B***: Normalized complexity ratings for circle, hexagon, square, and triangle SGP stimuli averaged across participants (colored dots), and across participants and shapes (black dots) versus the *c*-values. Error bars show standard error of mean.

Along with a subjective measure of complexity, i.e., the complexity ratings, Kolmogorov complexity approximated by the compressed file sizes of the images was used as an objective measure of stimulus complexity (Donderi & McFadden, 2005; Donderi, 2006). Kolmogorov complexity is defined as the length of the shortest program that can describe an output (Solomonoff, 1986). Kolmogorov complexity approximated by the compressed file size has been previously utilized in several other studies (Donderi & McFadden, 2005; Forsythe et al., 2008; Forsythe et al., 2011; Marin & Leder, 2013). In this study, PNG compression, which is a data format supporting lossless data compression was used. Furthermore, for comparison purposes correlational results of zip compression, which is often used in literature (Forsythe et al., 2011; Landwehr, Labroo & Herrmann, 2011; Marin & Leder, 2013), are also presented (Table 2.1). The PNG compression

and the zip compression performed very similarly, and were almost perfectly correlated (Pearson product-moment correlation coefficient $r = 0.999$).

Table 2.1: Bivariate correlations between various complexity measures, liking ratings, and the *c*-values. Reported values are *r*-values, i.e. Pearson product-moment correlation coefficients.* indicates p-values less than 0.01, and ** indicates *p*-values less than 0.001.

Measure	Normalized complexity ratings	File size (PNG)	File size (ZIP)	*c*-value	Normalized liking ratings
Normalized complexity ratings	1	0.946**	0.940**	-0.855**	-0.461*
File Size (PNG)	0.946**	1	0.999**	-0.863**	-0.615**
File Size (ZIP)	0.940**	0.999**	1	-0.865**	-0.628**
***c*-value**	-0.855**	-0.863**	-0.865**	1	0.705**
Normalized liking ratings	-0.461*	-0.615**	-0.628**	0.705**	1

For each participant the normalized liking ratings of circle, hexagon, square, and triangle SGP stimuli having the same *c*-value were averaged, resulting in 36 liking ratings per participant. The same procedure was repeated for the complexity ratings, resulting in 36 complexity ratings per participant. Figure 2.3 shows the normalized liking versus complexity ratings averaged across participants. The graph of liking ratings versus the complexity ratings revealed an inverted U-curve. Figure 2.4 shows the normalized liking and complexity ratings for each stimulus averaged across all shape types and participants versus the PNG compressed file sizes of the stimuli averaged across all shape types. Similar to the relationship between liking and subjective complexity ratings presented in Figure 2.3, liking ratings and the PNG compressed file sizes displayed an inverted U-curve relationship (Figure 2.4*A*). Furthermore, the average liking ratings and the file sizes were significantly negatively correlated (see Figure 2.4*A* and Table 2.1). On the other hand, as expected, complexity ratings increased with increased file size (Figure 2.4*B*). The two measures of complexity (i.e., the subjective ratings and the objective file sizes) were significantly and highly correlated (Table 2.1).

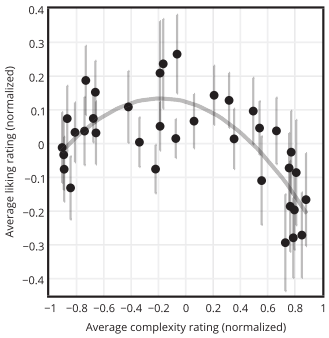

Figure 2.3: Liking versus complexity ratings of all participants. All ratings are normalized averages across all shapes and participants. Dots represent the stimuli, and the line represents a quadratic function fit. Error bars show standard error of mean. An inverted U-curve relationship between liking and complexity is visible.

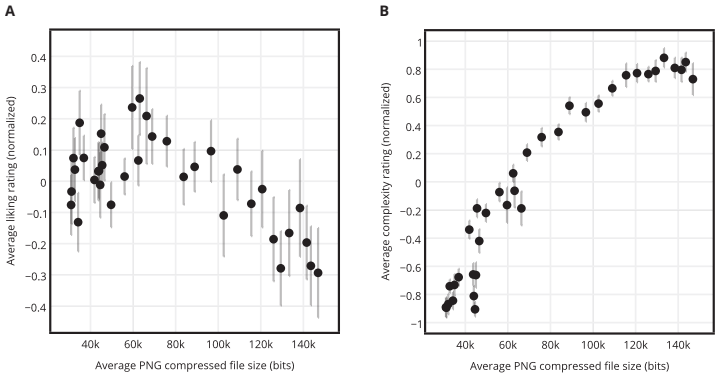

Figure 2.4: Liking and complexity ratings versus the PNG compressed file sizes. ***A***: Normalized average liking ratings for circle, hexagon, square, and triangle SGP stimuli averaged across shapes and participants, versus the average PNG compressed file sizes of the stimuli. ***B***: Normalized average complexity ratings for circle, hexagon, square, and triangle SGP stimuli averaged across shapes and participants, versus the average PNG compressed file sizes of the stimuli. Error bars show standard error of mean.

Next, a clustering approach was adopted in order to investigate the role of individual differences in forming this relationship by

identifying subgroups of individuals with different complexity preferences. As described earlier, normalized liking ratings of circle, hexagon, square, and triangle SGP stimuli having the same *c*-value were averaged, resulting in 36 liking ratings per participant. These liking ratings formed the input to the clustering algorithm. That is, the input to the clustering algorithm was a 30 by 36 matrix, containing the 36 normalized liking ratings for each one of the 30 participants. Using k-means clustering algorithm, several groupings of participants (k = 2, 3, 4, 5, 6, 7, 8) were identified. Specifically, participants were clustered into 2, 3, 4, 5, 6, 7, and 8 different subgroups based on their liking ratings averaged across stimuli having the same *c*-value. Afterward, to determine the optimal number of clusters among the formed subgroups, average silhouette values of the identified clusters were calculated.

Comparison of average silhouette values showed that the clusters with k = 2 had a significantly larger average silhouette value (Figure 2.5*A*) than all other values of k (Student's t-test, p extless 0.05 for all comparisons, Bonferroni corrected for multiple comparisons), meaning that for this dataset, the most appropriate grouping of the participants could be obtained when they were clustered into just two groups (Rousseeuw, 1987). Therefore, further analyses were performed only on the clusters obtained with k = 2. Figure 2.5*B* shows the silhouette values of participants as they were assigned to Cluster 1 and Cluster 2. Cluster 1 consisted of 20 participants (average age 25.1 ± 3.3, 7 males and 13 females), and Cluster 2 consisted of 10 participants (average age 22.7 ± 3.5, four males and six females).

Plotting the average liking versus complexity ratings of these two clusters separately revealed a negative relationship between liking and complexity ratings for Cluster 1 (Pearson product-moment correlation coefficient r = -0.789, $p < 0.001$), and a positive relationship between liking and complexity ratings for Cluster 2 (Pearson product-moment correlation coefficient r = 0.874,

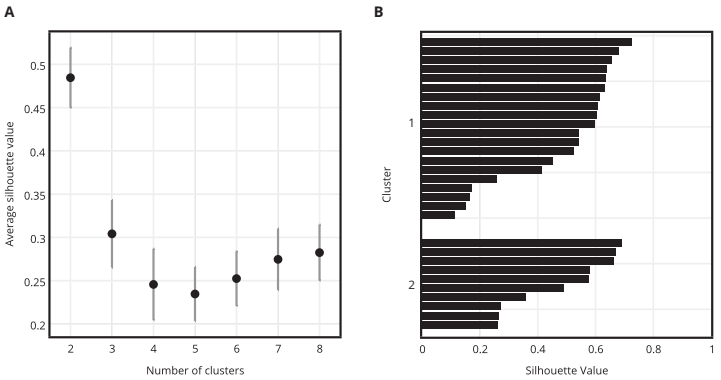

Figure 2.5: Evaluation of grouping of participants into different numbers of clusters. ***A***: Average silhouette values of clusters for k = 2, 3, 4, 5, 6, 7, and 8. Error bars show standard error of mean. k = 2, where the participants were divided into two clusters, had a significantly higher average silhouette value than the remaining numbers of clusters. ***B***: Silhouette value distribution of participants as they were assigned to Cluster 1 and Cluster 2, where k = 2.

$p < 0.001$). See Figure 2.6*A* for the results. In other words, participants in Cluster 1 liked the simple stimulus images more than the complex ones and participants in Cluster 2 liked the complex stimulus images more than the simple ones.

The same patterns were found for liking versus the PNG compressed file sizes (Figure 2.6*B*). Concretely, average liking ratings of participants in Cluster 1 and PNG compressed file sizes of the stimuli were highly negatively correlated (Pearson product-moment correlation coefficient r = -0.849, $p < 0.001$), whereas average liking ratings of participants in Cluster 2 and PNG compressed file sizes of the stimuli were highly positively correlated (Pearson product-moment correlation coefficient r = 0.842, $p < 0.001$). On the other hand, average complexity ratings of the participants in the two clusters were very similar (Figure 2.6*C*). The average complexity rating vs. the PNG compressed file size regression slopes of the two clusters did not differ significantly [ANCOVA, $F(1,68) = 3.07$, $p = 0.085$], suggesting that the per-

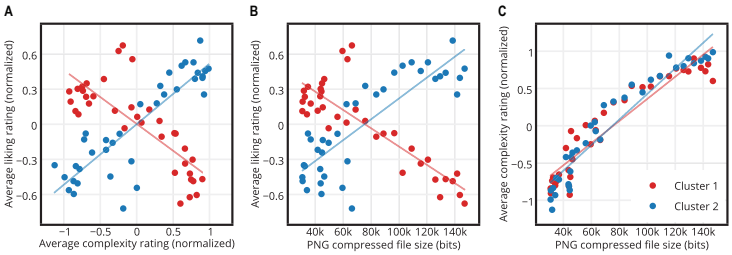

Figure 2.6: Ratings of the participants in the two clusters. ***A***: Normalized average liking versus complexity ratings of participants in Clusters 1 and 2. While participants in Cluster 1 liked the stimuli more as the stimulus complexity decreased, participants in Cluster 2 liked the more complex stimuli more. The dots represent the stimuli and the lines represent the regression lines. ***B***: Normalized average liking ratings of participants in Clusters 1 and 2 versus Kolmogorov complexity approximated by PNG compressed file sizes of the stimuli. The opposite patterns of the two clusters observed for liking versus subjective complexity in panel A are also observed here for liking versus Kolmogorov complexity. ***C***: Normalized average complexity ratings of the participants in Cluster 1 and Cluster 2 versus Kolmogorov complexity approximated by PNG compressed file sizes of the stimuli. There are no discernible differences between the average complexity ratings of the participants in different clusters.

ceived complexities of the stimuli did not differ between the two clusters.

Additionally, in order to statistically compare whether a single inverted U-curve pattern (i.e., a quadratic function) or a combination of two different preference patterns belonging to the two clusters of participants (i.e., combination of two linear functions) better explain the liking-complexity relationship in our data, we performed regression analyses. Specifically we compared the following two generalized linear mixed models:

$$\text{Liking} = \beta_0 + \beta_1 \text{Complexity} + \beta_2 \text{Complexity}^2 + b_0 \text{Participant} + \epsilon \tag{2.1}$$

$$\text{Liking} = \beta_0 + \beta_1 \text{Complexity} + \beta_2 \text{Cluster} + \beta_3 \text{Complexity} \times \text{Cluster} + b_0 \text{Participant} + \epsilon \tag{2.2}$$

Where β_i denotes the fixed-effect coefficients, b_0 denotes the random-effect coefficients and ϵ denotes the residuals. Random-effects terms were included in both models in order to account for the repeated measures structure of the data. Models were implemented using MATLAB. Since liking ratings were observed to be normally distributed, normally distributed responses and identity link function options were selected during the implementation. Table 2.2 shows the estimated coefficients of the two models.

Table 2.2: Estimated fixed-effect coefficients of the quadratic and cluster-based models. The meanings of the abbreviations in the column headers are as follows: SE = standard error, DF = degrees of freedom and CI = confidence interval.

Model	Coefficient name	Estimate	SE	DF	{t}-statistic	{p}-value	Lower CI (95%)	Upper CI (95%)
Fixed-effect coefficients of the quadratic model	Intercept	0.067	0.026	1077	2.530	0.012	0.015	0.119
	Complexity	-0.045	0.024	1077	-1.912	0.056	-0.092	0.001
	Complexity^{2}	-0.117	0.034	1077	-3.425	<0.001	-0.185	-0.050
Fixed-effect coefficients of the cluster-based model	Intercept	~0	0.048	1076	~0	1	-0.095	0.095
	Complexity	-1.004	0.064	1076	-15.662	<0.001	-1.130	-0.878
	Cluster	~0	0.034	1076	~0	1	-0.067	0.067
	Complexity * Cluster	0.719	0.045	1076	15.978	<0.001	0.631	0.807

Next, the two models were compared using a simulated likelihood ratio test with 1000 simulations. Table 2.3 shows the results of simulated likelihood ratio test. The cluster-based model had lower Akaike information criterion (AIC) and Bayesian information criterion (BIC) values than the quadratic model, indicating that the cluster-based model is the better fitting model (Hox, 2002). Note that the *p*-value for the simulated likelihood ratio test was less than 0.001, further demonstrating that the cluster based model significantly better explains the data.

Table 2.3: Simulated likelihood ratio test results. The meanings of the abbreviations in the column headers are as follows: DF = degrees of freedom (number of free parameters in the model), AIC = Akaike information criterion for the model, BIC = Bayesian information criterion for the model, Log Likelihood = maximized log likelihood for the model, LRT-statistic = likelihood ratio test statistic and CI = confidence interval.

Model	DF	AIC	BIC	Log likelihood	LRT-statistic	*p*-value (95% CIs)
Quadratic model	5	1923.5	1948.5	-956.77	217.49	< 0.001
Cluster-based model	6	1708	1737.9	-848.02		(0.00002 - 0.006)

As a final analysis, we looked at the response times of the participants. Participants in Cluster 1, who had a preference toward simpler patterns, had significantly shorter average response times ($M = 1.639$ s, $SD = 0.108$ s) for liking ratings than participants in Cluster 2 ($M = 1.803$ s, $SD = 0.197$), who preferred more complex patterns (Student's t-test, $p < 0.001$).The average response times for complexity ratings of participants in Cluster 1 ($M = 1.477$ s, $SD = 0.103$ s) and Cluster 2 ($M = 1.494$ s, $SD = 0.210$) did not differ (Student's t-test, $p = 0.396$).

2.4 Discussion

In this study, abstract computer-generated art of varying levels of complexity was evaluated in terms of liking and complexity. Consistent with the literature, we found an inverted U-curve relationship between liking and complexity of these images. Next, utilizing a data-driven clustering approach, we revealed subgroups of people with different preferences for image complexity. Note that we neither had an *a priori* assumption regarding the number of clusters nor the shape of the liking-complexity relationships in the different clusters. Following the clustering approach, two subgroups having opposite preferences for image complexity (one which liked simple patterns more than the complex ones, and another which liked complex patterns more than the simple ones)

were identified. The two groups differed in terms of how much they liked the complex or simple patterns, but not how they perceived complexity. Furthermore, a comparison of a quadratic model (representing the inverted U-curve relationship) and a cluster-based model (representing the combination of two linear relationships) revealed that the cluster-based model was a better fit to the data at hand. Interestingly, the group of participants who liked the simpler patterns more were faster in their liking evaluations compared to the group that preferred complex patterns. In contrast, there were no differences between the groups in terms of the time they took to evaluate the complexity of the images.

Possible Interpretations of the Results

An important point to keep in mind when attempting to interpret our results is the exploratory – rather than hypothesis driven – approach that we took in this study. One implication of this approach is that we can describe the differences between the two identified clusters in terms of the measurements at hand, but cannot make claims about causal relationships regarding the factors influencing these results. For example, by clustering participants based on their liking ratings, a non-random assignment of participants to one of the two groups was introduced. This in turn might have increased the possibility of groups having different distributions in terms of several unidentified factors such as intelligence, personality, art experience, motivation, etc. Because one or multiple of these unidentified factors might have had an influence on the response time results, the strength of conclusions that can be derived from the results about the response time differences becomes limited. In other words, our interpretations of the results should be viewed in light of the fact that all the analyses and statements trying to characterize the clusters are *post hoc*.

We showed that the group of participants who preferred simpler patterns were faster in their liking evaluations compared to the

group that preferred complex patterns. The liking rating response time differences that we identified between the groups were not present for the complexity rating responses. This leads us to think that the unidentified factors contributing to these response time differences should be more specifically related to aesthetic evaluation of images (e.g., art experience/education, art interest, motivation, personality), rather than a more general factor which would also effect the complexity evaluation response times (e.g., intelligence). Furthermore, on average, the complexity of the images were perceived similarly in the two clusters as shown in Figure 2.6*C*. This suggests that the preference differences of the two clusters cannot be explained by the differences in their perception of complexity.

A partial explanation to our results could be provided by the fluency theory (Reber, Schwarz & Winkielman, 2004). According to the fluency theory, experience of fluent processing results in positive affect toward stimuli. Based on this theory, one would expect decreased liking toward complex (and hence less fluently processed) stimuli compared to simpler (and hence more fluently processed) stimuli. The majority of the participants in our experiment (20 out of 30) were assigned to Cluster 1 which showed a monotonic decrease in liking for stimuli with increased complexity. Their behavior is in line with the fluency theory. However, the average tendency of the remaining ten participants in Cluster 2 was a monotonic increase in liking for stimuli with increased complexity, which is difficult to account for with the fluency theory of aesthetic liking. Future studies can investigate differential effects of fluency on participants who have different complexity preferences in order to evaluate the merit of such an interpretation.

A recent framework by Graf and Landwehr (2015), called the Pleasure-Interest Model of Aaesthetic Liking (PIA Model), claims to provide a better explanation for contradictory preference patterns for aesthetic stimuli that are easy or difficult to process. According to Graf and Landwehr (2015), an aesthetic object may

be processed in two stages. First an automatic processing takes place, and then if the viewer is motivated enough to process the stimuli further, a controlled processing follows. Similar to the fluency theory, the PIA Model predicts that merely automatic processing of stimuli would result in a monotonic decrease of liking as the stimulus complexity increases. The model further predicts that controlled processing could result in an inverted U-curve if the complexity levels of stimuli are high enough to cause dislike and confusion. Our results do not conflict with this model, however to explain the results in terms of the PIA model, assumptions need to be made. These assumption are related to the motivation levels of the participants, perceived complexity of the stimulus material and factors affecting the response times. Testing these assumptions would go beyond the scope of this paper. Additional studies would be required to test the PIA model.

Relation to Past and Future Research

We believe that it is important to try and characterize the groups of people with different complexity preferences as found in the present study. Previous studies give some valuable insight in this direction. For example, preference for complexity has long been associated with creativity and artistic tendencies (Barron & Welsh, 1952; Mackinnon, 1962; Barron, 2012; McWhinnie, 1968). In fact, the Barron-Welsh art scale (Barron & Welsh, 1952) which measures an individual's preference for complexity has been used to assess creativity in several previous studies (for a review, see (Gough, Hall & Bradley, 1996). Along with creativity and artistic tendencies, a person's age has been shown to affect their preference for complexity. Particularly, older individuals have been shown to prefer simpler visual stimuli (Munsinger et al., 1964; Alpaugh & Birren, 1977; Crosson & Robertson-Tchabo, 1983). In the present study, although the average age of Cluster 1 (M = 25.1, SD = 3.3) was higher than that of Cluster 2 (M = 22.7, SD = 3.5), this difference was very small and not significant (Student's

t-test, $p = 0.082$). However, since our sample of participants consisted of a similarly aged group of young adults between the ages of 20 and 32, such a result was expected.

More recently, the personality traits 'openness to experience' and 'conscientiousness' (of the Big Five personality inventory by Goldberg (1999) along with 'frequency of visits to galleries/museums' have been shown to correlate with preference for more complex art (Chamorro-Premuzic et al., 2009; Cleridou & Furnham, 2014). However, it is important to note that these correlations were only able to account for a limited amount of the variability in the data. Nevertheless, it is important for future studies to investigate the distribution of personality traits as well as artistic and creative tendencies within the clusters of participants who have different complexity preferences.

Here, we focused on complexity because it is a relatively general and overarching concept that has been shown to relate to the appreciation of a stimulus on several dimensions such as the number of elements, irregularity of shape and arrangement, heterogeneity of elements and asymmetry (Berlyne, 1971). Additionally, in the past decades various measures of pattern complexity have been suggested, developed and shown to be relevant for perception, e.g., Birkhoff's complexity elements, C (Birkhoff, 1933; Eysenck, 1941), the amount of structural information (Leeuwenberg, 1969; Leeuwenberg, 1971; Boselie, 1984; van der Helm, van Lier & Leeuwenberg, 1992), amount of algorithmic information, i.e., Kolmogorov complexity (Donderi & McFadden, 2005; Marin & Leder, 2013), and the amount of self-similarity, i.e., fractal dimension (Mandelbrot, 1977), (Mandelbrot, 1977; Cutting & Garvin, 1987; Spehar et al., 2003). Therefore, we found complexity to be a good starting point for studying the individual differences in preferences. It would however be interesting to further investigate how the individual differences in complexity preferences relate to the previously identified individual differences in preferences

for other visual perceptual attributes, such as harmony (Palmer & Griscom, 2012) or symmetry (Jacobsen & Höfel, 2002).

Recently, Spehar et al. (2015) have shown that participants' visual sensitivity to various visual properties (e.g., amplitude spectrum characteristics of synthetic images and spatial frequency of sine-wave gratings) highly correlate with their visual preferences. In a future study, it would be very interesting to investigate the relationship between visual sensitivity of individuals and their preference for complexity. Furthermore, another interesting question to investigate would be whether or not a clustering of participants as found in our study would be observed for more complex artistic stimuli with interpretable contents.

Conclusion

The results of our study have both theoretical and practical implications. Here, we showed that one of the most well-known rules of aesthetic preference, i.e., the inverted U-curve of preference for complexity, can in fact be an artifact that arises from selecting a non-ideal analysis method. By employing an averaging approach, most experimental aesthetics studies risk reaching conclusions about a non-existent average observer. Besides the need for utilization of new analysis methods that take into account the differences between individuals, theoretical implications of this finding include a need for re-evaluation of established rules of human aesthetic preferences and revision of existing theories, in a way that would explain e.g., the monotonically increasing and monotonically decreasing liking as a function of complexity, rather than a global mid-level complexity preference. Similarly, in the practical sense, our results are relevant for designers and artists. Rather than opting for a mid-level complexity in their designs to please the average observer, they can utilize more targeted design strategies for appealing to different groups of individuals. However, characteristics of these groups remain to be identified.

Supplementary Materials

Stimulus Set

The stimulus set consisted of 144 grayscale square statistical geometric pattern (SGP) images with a side length of 800 pixels. The images were generated using the algorithm developed by Shier (2011). Specifically, there were 36 circle, 36 hexagon, 36 square and 36 triangle SGPs with equally spaced c-parameter values ranging from 0.2 to 1.7. Please see below for the resized versions of the complete set of stimulus images.

Algorithm for Generating SGP Images

The initial steps for generating an SGP image (as used in our stimulus set) include defining an empty enclosed surface to be filled, selecting a shape which will be used to fill this surface, defining stopping criteria and selecting a value for the algorithm parameter *c*. The algorithm then starts by placing the first shape at a random coordinate inside the defined enclosed surface. At the next iteration, after an intersection check ensuring that no two shapes overlap, a smaller shape is placed at another random location. The size of the enclosed surface, the value of the algorithm parameter c, and the set maximum number of iterations together determine the size of the first shape to be placed on the surface. Furthermore, the value of parameter *c* also determines the speed with which the shape area decreases at each iteration. Concretely, at each iteration, the shape area decreases fast for large values of *c*, whereas it decreases slowly for small values of *c* (see Figures 2.7-2.10).

A collection of resources regarding the algorithm for generating SGP images can be found online at John Shier's website http://john-art.com/stat_geom_linkpage.html. Furthermore, Paul

Bourke has a very informative website on the topic: http://paulbourke.net/texture_colour/randomtile/.

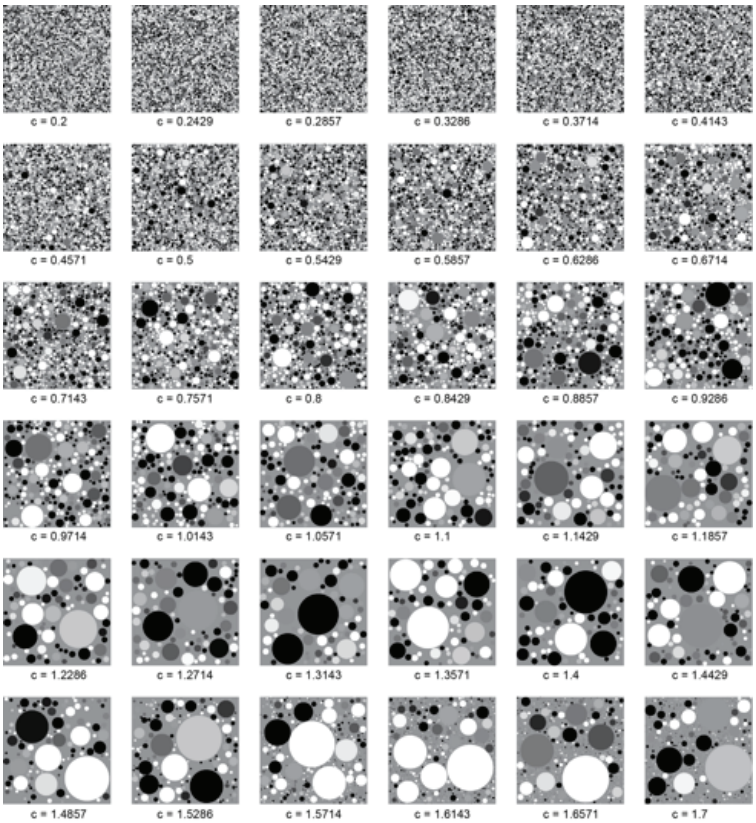

Figure 2.7: Circle SGP images. Equally spaced c-parameter values range from 0.2 to 1.7.

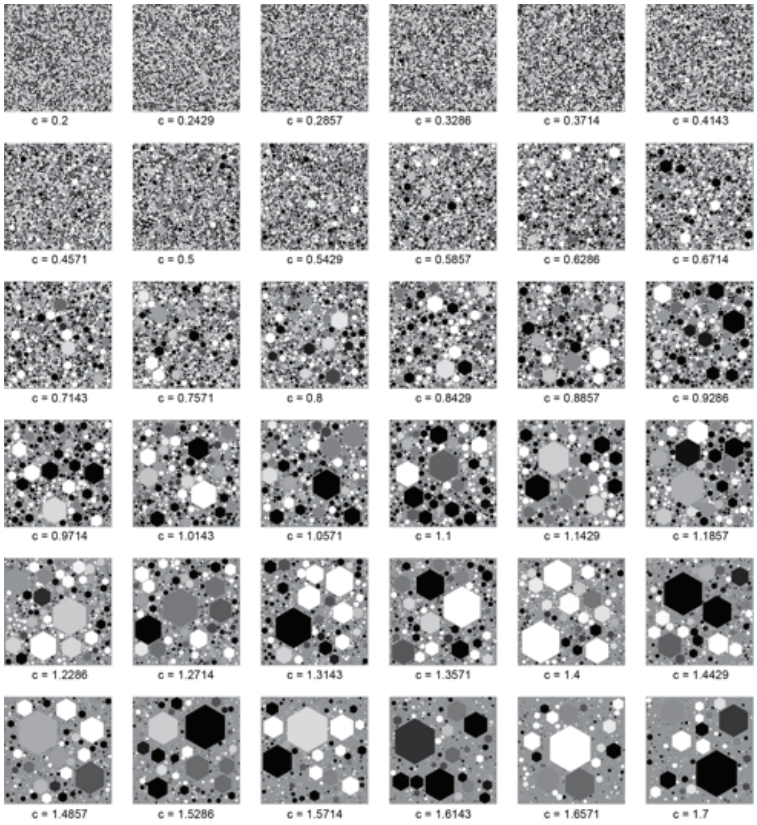

Figure 2.8: Hexagon SGP images. Equally spaced c-parameter values range from 0.2 to 1.7.

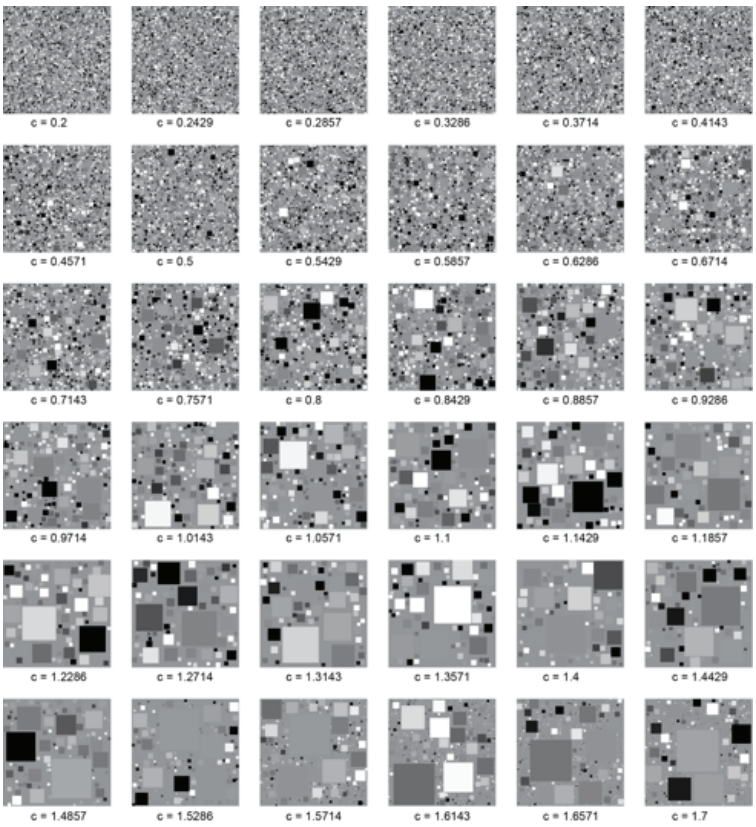

Figure 2.9: Square SGP images. Equally spaced c-parameter values range from 0.2 to 1.7.

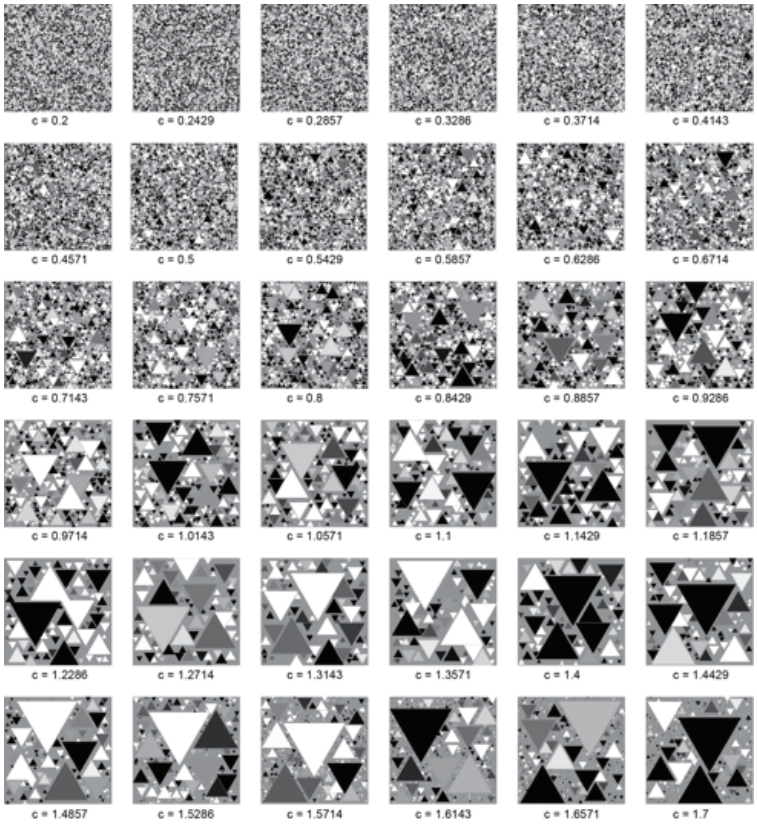

Figure 2.10: Triangle SGP images. Equally spaced c-parameter values range from 0.2 to 1.7.

3

Decomposing Complexity Preferences for Music

This chapter is based on Güçlütürk, Y., and van Lier, R. (2018). Decomposing complexity preferences for music. PsyArXiv.

3.1 Introduction

Like its visual counterpart, music complexity has been shown to have an important role in determining whether a song will be liked or not (Heyduk, 1975; Marin & Leder, 2016, 2013; North & Hargreaves, 1995; Orr & Ohlsson, 2001). In the context of music, complexity is a term that encompasses several acoustic and perceptual properties of music, such as tempo, predictability, variety of instruments, harmony, rhythm, etc. As suggested by Berlyne (1971), many music studies reported an inverted U-curve relationship between stimulus complexity and liking (North & Hargreaves, 1995; Orr & Ohlsson, 2001). However other relationships between liking and complexity of music have also been frequently observed (Marin & Leder, 2016, 2013; Orr, 2005).

Besides complexity, factors such as familiarity, emotional valence, music genre (which also relates to complexity), etc. have been known as important modulators of music preferences (for a review see (Corrigall & Schellenberg, 2015)). In fact, recently Madison and Schiölde (2017) showed that familiarity increases liking independent of the complexity of the music excerpts. Such results highlight the importance of stimulus selection in studies that aim to investigate liking and preferences with ecologically valid stimuli. The songs used in the current study were previously used in several other studies investigating music preferences with robust results (Greenberg et al., 2015; Rentfrow et al., 2012; Rentfrow et al., 2011) and were selected to span a variety of music types and the five factors of the MUSIC model (Rentfrow et al., 2011) allowing us to make more general conclusions rather than e.g., genre specific ones. Furthermore, the songs were selected among not well-known songs in Getty Images to avoid familiarity effects (Rentfrow et al., 2011).

Recently, we demonstrated complexity as a major factor for explaining individual differences in visual preferences for abstract

digital art (Güçlütürk et al., 2016b). We showed that participants could best be separated into two groups based on their liking ratings for abstract digital art comprising geometric patterns: one group with a preference for complex visual patterns and another group with a preference for simple visual patterns. These two opposite complexity preferences emerged from what initially appeared to be an inverted U-curve when the data were simply averaged. These preference relationships were obtained for a highly-controlled set of visual stimuli, varying in only a few perceptual dimensions such as the number and size of elements.

Our findings were later replicated in the visual domain by identification of robust preference differences for fractal-like images (Spehar, Walker & Taylor, 2016), where four different preference patterns, including a group with linearly increasing, another group with linearly decreasing, another intermediate group with mid-level spectral slope preference and finally a group without any particular preference, were identified. However, the stimuli sets used in this study were also strictly controlled grayscale images, therefore it is still not known whether these results would further generalize to more ecologically valid stimuli and/or other sensory modalities.

In the present study, we extend our investigations for complexity preferences from highly controlled visual stimuli to ecologically valid stimuli in the auditory modality. Additionally, to gain further insights about participant characteristics who have different complexity preferences, we investigate various demographic and personality measures. We test whether grouping individuals based on differences in their liking ratings for song excerpts would result in a grouping which also reflects distinct complexity preferences. Similar to previous results in visual preferences (Güçlütürk et al., 2016b), we show that music preferences are highly influenced by stimulus complexity. Using k-means clustering algorithm, we demonstrate this by clustering a large number of participants based on their liking ratings for song excerpts from various musical

genres. Our results show that, based on their liking ratings, participants can again best be separated into two groups: one group with a preference for more complex songs and another group with a preference for simpler songs. Furthermore, we present characteristics of individuals with different music complexity preferences in terms of demographic and personality measures as well as estimated autism spectrum quotient (AQ), providing insights into factors related to complexity preferences for music.

3.2 Materials and Methods

Data

The data were derived from 'Study 2' of Greenberg et al. (2015). This dataset consists of liking ratings by 353 participants for 25 different song excerpts as well as demographic information for the participants, such as age, gender, occupation, etc. and responses to two personality questionnaires measuring i) emotional quotient (EQ) and ii) (revised) systemizing quotient (SQ-R). For details regarding the data collection, we refer the reader to Greenberg et al. (2015). Below we briefly describe the relevant aspects.

Participants and Procedures

Participants were Amazon Mechanical Turk workers who participated by filling an online survey that was hosted by Qualtrics. Total number of participants who completed all the required measures was 353. Among these 353 participants, 220 (62%) were female and 133 (38%) were male. The ages of the participants were between 18 and 68 ($M = 34.10$, $SD = 12.27$). This research was given ethical approval by the Psychology Research Ethics Committee of the University of Cambridge in August 2013.

Stimuli

Stimulus material consisted of 15-second-long 25 song excerpts that were previously used in several music preference studies (Greenberg et al., 2015; Rentfrow et al., 2012; Rentfrow et al., 2011). The songs were selected by Rentfrow et al. (2011) to span the five broad dimensions of the MUSIC model, which is a model for explaining individual differences in music preferences. MUSIC is an acronym for the following: Mellow (romantic, relaxing, unaggressive, sad, slow, and quiet music; example genres: soft rock, R&B, and adult contemporary); Unpretentious (uncomplicated, relaxing, unaggressive, soft, and acoustic music; example genres: country, folk); Sophisticated (inspiring, intelligent, complex, and dynamic music; example genres: classical, operatic, avant-garde, world beat, and traditional jazz); Intense (distorted, loud, aggressive, and not relaxing, romantic, nor inspiring music; example genres: classic rock, punk, heavy metal, and power pop); and Contemporary (percussive, electric, and not sad music; example genres: rap, electronica, Latin, acid jazz, and Euro pop).

The genres of the 25 stimulus songs were as follows: rock-n-roll, adult contemporary, electronica, soft rock, europop, R&B soul, rap, avant-garde classical, classical, latin, traditional jazz, world beat, classic rock, heavy metal, punk, bluegrass, mainstream country, new country. Since different genres tend to have different instrumental complexity levels (Percino et al., 2014), this variety serves to diversify the levels of complexity in the dataset. A full list of the songs in order of increasing complexity is available in supplementary materials.

Measures

In the current study, as part of the demographic measures we used the age and gender variables. Other measures used in the study are described below.

Empathy Quotient Empathy was measured using the 60-item self-report EQ questionnaire which measures the affective and cognitive components of empathy in adults (Baron-Cohen & Wheelwright, 2004). Empathy as measured by this questionnaire was defined as follows: "*the drive or ability to attribute mental states to another person/animal, and entails an appropriate affective response in the observer to the other person's mental state.*" Each statement was evaluated by the participants on a four-point scale consisting of: strongly disagree, slightly disagree, slightly agree, or strongly agree.

Systemizing Quotient-Revised Systemizing was measured using the 75-item SQ-R questionnaire (Wheelwright et al., 2006), and it was defined as follows: "*the drive to analyze, understand, predict, control and construct rule-based systems.*" Similar to EQ, each statement in the questionnaire was evaluated by the participants on a four-point scale consisting of: strongly disagree, slightly disagree, slightly agree, or strongly agree.

Brain Types Five cognitive 'brain types' as per the E-S theory (Baron-Cohen, Richler, Bisarya, Gurunathan & Wheelwright, 2003) were calculated for each participant based on the standardized differences between EQ and SQ measures. The five brain types are defined as follows:

- Type E: Individuals who have more developed empathizing drive/abilities than systemizing ones

- Type B: Individuals who have equally developed empathizing and systemizing drive/abilities

- Type S: Individuals who have more developed systemizing drive/abilities than empathizing ones

- Extreme type E: Individuals who have normal or overdeveloped empathizing drive/abilities and underdeveloped systemizing ones (There were no individuals with this brain type in the current sample)

- Extreme type S: Individuals who have normal or overdeveloped systemizing drive/abilities and underdeveloped empathizing ones

Autism Spectrum Quotient AQ measures where an individual lies in the continuum of autistic traits (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001). Autism spectrum conditions usually manifest themselves with difficulties in empathy and an increased tendency for systemizing behavior (Wheelwright et al., 2006). In this study, since the original dataset did not contain AQ measurements, AQ for each participant was estimated using their EQ and SQ-R scores as described by Wheelwright et al. (2006). Specifically, AQ was calculated using the following two formulas:

$$\mathrm{AQ_m} = 0.089\mathrm{SQ\text{-}R} - 0.25\mathrm{EQ} + 21.6 \tag{3.1}$$

$$\mathrm{AQ_f} = 0.089\mathrm{SQ\text{-}R} - 0.25\mathrm{EQ} + 22.7 \tag{3.2}$$

for males ($\mathrm{AQ_m}$) and females ($\mathrm{AQ_f}$), respectively. In general population, males and females are known to differ in AQ, EQ, and SQ score distributions (Baron-Cohen, 2010; Wheelwright et al., 2006), as reflected by the above formulas.

Liking Ratings Each participant listened to all 25 excerpts and provided a rating of how much they liked each excerpt. The ratings were collected on a nine point Likert scale (1 = dislike extremely; 2 = dislike very much; 3 = dislike moderately; 4 = dislike slightly; 5 = neither like nor dislike; 6 = like slightly; 7 = like moderately; 8 = like very much; 9 = like extremely).

Complexity Ratings Complexity ratings for the song excerpts were previously collected as part of another study (Rentfrow et al., 2011). In the current study, complexity score of each song is taken as the average complexity rating given by the 40 'judges' that rated the songs in the original study.

Complexity Preference Complexity preference of each of the 353 participants were calculated as follows:

$$\mathrm{CP}^{(n)} = \sum_{i=1}^{25} \mathrm{L}_i^{(n)} \mathrm{C}_i \tag{3.3}$$

where $\mathrm{CP}^{(n)}$ is the complexity preference of the *n*th participant, $\mathrm{L}_i{}^{(n)}$ is the liking rating of the *n*th participant for the *i*th song and C_i is the complexity score of the *i*th song.

3.3 Results

Clusters

Participants were clustered into groups based on their liking ratings for the 25 songs. For this, k-means clustering algorithm (Lloyd, 1982) was used. To explain briefly, cluster analysis groups individual elements in a set such that the similarity of elements

assigned to the same cluster is maximized, whereas the similarity between different clusters is minimized. A distance measure between the elements is often used as a measure of similarity. K-means clustering algorithm iteratively partitions data into k clusters by assigning each data point to a cluster such that the within cluster sum of squares is minimized. This is equivalent to assigning each data point to the nearest cluster.

Here, k-means clustering algorithm was used for clustering the 353 participants into subsamples such that participants with similar song preferences would be assigned to the same cluster. Specifically, we clustered the participants based on their liking ratings of the 25 songs. We tested several numbers of clusters ($k = [2, 10]$) to determine the best separation of the participants (Figure 3.1). After obtaining the clusters, to determine the optimum number of clusters, average silhouette values across all data were calculated for each value of k. The silhouette value measures a data point's within cluster similarity in comparison to its between cluster similarity (Rousseeuw, 1987). We found that when $k = 2$, the average silhouette value was significantly larger than the remaining values of k (Student's t-test, $p < 0.001$ for all comparisons, Bonferroni corrected for multiple comparisons). Since a high average silhouette value indicates a better clustering, we concluded that clustering participants into two groups resulted in the best separation and the most appropriate grouping of the participants (Rousseeuw, 1987). Therefore, further analyses were performed on these two clusters. Specifically, in Cluster 1 there were 190 participants (103 females and 87 males, average age $\pm$ SD = 31.36 $\pm$ 10.24) and in Cluster 2 there were 163 participants (117 females and 46 males, average age $\pm$ SD = 37.31 $\pm$ 13.63).

Additionally, we calculated the intraclass correlation (ICC) indices for the whole sample of participants as well as for each cluster to evaluate the agreement of participants within these groups (Shrout & Fleiss, 1979). Specifically, we calculated ICC(2,

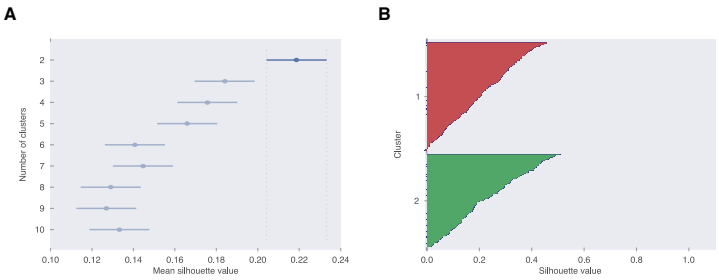

Figure 3.1: Evaluation of clusters. ***A***: Comparison of average silhouette values of clustering the participants with $k = [2, 10]$, where $k = 2$ appears to be a significantly better grouping of the 353 participants compared to the remaining values. ***B***: Silhouette value distribution of participants as they were assigned to Cluster 1 and Cluster 2, where $k = 2$.

1) i.e. two-way random single measures for all the participants together, for participants in Cluster 1 alone, and for those in Cluster 2 alone. ICC(2, 1) for all participants was 0.10, whereas after clustering the participants it increased to 0.20 and 0.24 for Cluster 1 and Cluster 2, respectively. Clustering the participants significantly increased the ICC indices (F-test, $p < 0.001$). This analysis confirms that clustering the participants allowed us to obtain subsets of the sample that agreed more with each other.

We then calculated the average normalized liking ratings of all participants and plotted these average ratings vs. normalized complexity scores of the 25 song excerpts. We did the same for the ratings of the participants who were assigned to Cluster 1 and Cluster 2 separately as well (Figure 3.2). Visual inspection of these liking vs. complexity graphs of the two clusters revealed clearly distinct relationships between the liking and complexity variables. While the 190 participants in Cluster 1 on average had a clear preference for more complex songs, the 163 participants in Cluster 2 had an opposite average preference pattern of a preference for simpler songs.

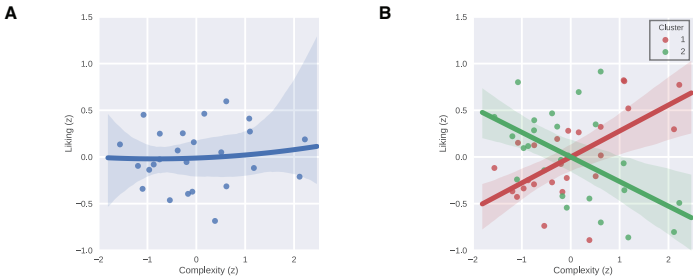

Figure 3.2: Liking ratings of the participants as a function of stimulus complexity. ***A***: Average normalized liking ratings of all participants vs. normalized complexity scores of the 25 song excerpts. Dots represent the stimuli, and the line represents a quadratic function fit. Shaded area indicates error of fit. ***B***: Average normalized liking ratings of participants in Cluster 1 and Cluster 2 separately vs. normalized complexity scores of the 25 song excerpts. Dots represent the stimuli, and the two lines represent the regression lines. Shaded area indicates error of fit.

Next, to statistically compare a single quadratic function and a combination of two linear functions in terms of how much they explain the liking-complexity relationship in the data, we performed regression analyses. We compared the following two generalized linear mixed models:

$$\text{Liking} = \beta_0 + \beta_1 \text{Complexity} + \beta_2 \text{Complexity}^2 + b_0 \text{Participant} + \epsilon \tag{3.4}$$

$$\text{Liking} = \beta_0 + \beta_1 \text{Complexity} + \beta_2 \text{Cluster} + \beta_3 \text{Complexity} \times \text{Cluster} + b_0 \text{Participant} + \epsilon \tag{3.5}$$

where β_i denotes the fixed-effect coefficients, b_0 denotes the random-effect coefficients and ϵ denotes the residuals. Random-effects terms were included in both models in order to account for the repeated measures structure of the data. Models were implemented using MATLAB. Since liking ratings were observed to be

normally distributed, normally distributed responses and identity link function options were selected during the implementation. Table 3.1 shows the estimated coefficients of the two models.

Table 3.1: Estimated fixed-effect coefficients of the quadratic and cluster-based models. Abbreviations: SE = standard error, DF = degrees of freedom and CI = 95% confidence interval.

Fixed-effect coefficients	Coefficient name	Estimate	SE	DF	*t*-statistic	*p*-value	Lower CI	Upper CI
Quadratic model	Intercept	-0.01	0.01	8822	-0.85	0.394	-0.04	0.02
	Complexity	0.02	0.01	8822	1.71	0.088	-0.00	0.04
	Complexity^2	0.01	0.01	8822	1.30	0.193	-0.01	0.03
Cluster-based model	Intercept	~0	0.03	8821	~0	1	-0.06	0.06
	Complexity	0.82	0.03	8821	25.95	~0	0.76	0.88
	Cluster	~0	0.02	8821	~0	1	-0.04	0.04
	Complexity × Cluster	-0.54	0.02	8821	-26.48	~0	-0.58	-0.50

Next, the two models were compared using a simulated likelihood ratio test with 1000 simulations. Table 3.2 shows the results of simulated likelihood ratio test. The cluster-based model had lower Akaike information criterion (AIC) value and Bayesian information criterion (BIC) value than the quadratic model, indicating that the cluster-based model is the better fitting model (Hox, Moerbeek & van de Schoot, 2010). Note that the *p*-value for the simulated likelihood ratio test was less than 0.001, further demonstrating that the cluster based model significantly better explains the data.

Table 3.2: Simulated likelihood ratio test results. Abbreviations: DF = degrees of freedom, AIC = Akaike information criterion for the model, BIC = Bayesian information criterion for the model, Log Likelihood = maximized log likelihood for the model, LRT-statistic = likelihood ratio test statistic and CI = 95% confidence interval.

Model	DF	AIC	BIC	Log likelihood	LRT-statistic	*p*-value (CI)
Quadratic	5	24610	24645	-12300	673.09	0.001
Cluster-based	6	23939	23981	-11963		(0.00003 - 0.006)

In summary, these analyses demonstrate that rather than averaging the data of all participants, grouping them by means of a clustering algorithm results in two groups with more homogenous

and distinct complexity preference patterns that better explain the data. Furthermore, these two opposite complexity preference patterns that we now show for music stimuli coincide with the earlier findings that were established in the visual modality (Güçlütürk et al., 2016b). Taken together, these results suggest that this complexity-liking relation is not restricted to a specific domain and that it could rather be a supramodal characteristic of sensory preferences.

Characterizing the Participants in the Different Clusters

Next, we performed a number of post-hoc analyses to investigate the characteristics of the people assigned to different clusters. As demographic measures, we looked at whether age and gender distribution differed between the two clusters. As personality measures, we looked at the distributions of EQ, SQ, 'brain type', and AQ, which we estimated using EQ and SQ scores of individuals. We will first focus on each of the demographic and personality measures, and next consider their mutual contribution to determine the relative influence of these characteristics (by applying a logistics regression analyses).

Table 3.2 shows the how these characteristics were among the whole sample and the clusters as well as the results of Holm-Bonferroni corrected statistical tests comparing the two clusters. We found that both demographic and personality measures were informative for characterizing the participants in different clusters (Figure 3.3 and Table 3.3). Particularly, we found that participants in Cluster 2 who preferred simpler music were significantly older than those in Cluster 1 who preferred more complex songs. In terms of the gender distribution, majority of the male participants were assigned to Cluster 1 (65% of the males were assigned to Cluster 1 and 35% to Cluster 2), whereas for females this pattern was reversed and the difference between the number

of female participants in the two clusters was smaller (47% of the females were assigned to Cluster 1 and 53% to Cluster 2). While tests showed that the gender and age had a significantly different distribution in the two clusters, the distribution of the personality predictors EQ and SQ did not differ. However, we found significant differences between the distribution of different brain types (especially type S and type B) in the two clusters. Specifically, those participants who were type S were more likely to be assigned to the cluster that preferred complex stimuli (i.e. Cluster 1), whereas the participants with type B brain types were more likely to prefer simple stimuli (i.e. be assigned to Cluster 2).

Table 3.3: Characteristics of the participants assigned to the two clusters. * indicates p-values < 0.05, ** indicates p-values < 0.001, and *** indicates p-values $<< 0.001$ after Holm-Bonferroni correction. [1] Note that complexity preference was normalized to have zero mean and unit variance. [2]Note that since the normalized complexity preference is defined as a function of liking ratings, it is expected to be significantly different in the two clusters.

	All mean ± SD or #participants	Cluster 1 mean ± SD or #participants	Cluster 2 mean ± SD or #participants	p-value	Test statistic (test)
Age	34.11 ± 12.27	31.36 ± 10.24	37.31 ± 13.63	4e-05**	$\chi^2 = 16.87$ (Kruskal-Wallis)
Gender	#F: 220 #M: 133	#F: 103 #M: 87	#F: 117 #M: 46	7e-04*	$\chi^2 = 11.53$ (χ^2)
EQ	41.62 ± 12.37	40.67 ± 12.62	42.74 ± 12.02	0.1176	$t = -1.57$ (Student's t)
SQ-R	63.90 ± 20.34	65.74 ± 20.84	61.74 ± 19.60	0.0655	$t = -1.85$ (Student's t)
Brain type	#E: 59 #B: 103 #S: 182 #XS: 9	#E: 26 #B: 45 #S: 113 #XS: 6	#E: 33 #B: 58 #S: 69 #XS: 3	0.0070*	$\chi^2 = 12.11$ (χ^2)
AQ_{all}	17.98 ± 2.98	18.38 ± 2.96	17.51 ± 2.94	0.0030*	$t = 2.77$ (Student's t)
AQ_f	17.19 ± 2.86	17.53 ± 2.97	16.88 ± 2.74	0.0475	$t = 1.68$ (Student's t)
AQ_m	19.30 ± 2.70	19.40 ± 2.63	19.11 ± 2.84	0.2805	$t = 1.68$ (Student's t)
CP	0 ± 1[1]	0.68 ± 0.68 (complex)	-0.79 ± 0.68 (simple)	~0***[2]	$t = 0.58$ (Student's t)

Based on the literature that identified enhanced sensory processing in autism spectrum condition (ASC) (Dakin & Frith, 2005; Haesen et al., 2011) in combination with the fluency theory of aesthetic pleasure (Reber et al., 2004), we hypothesized that the

participants who were assigned to the cluster with high complexity preference, i.e. Cluster 1, would on average have higher AQ levels. As expected, one-tailed Student's *t*-tests comparing the two clusters revealed that the average AQ was significantly higher for the participants of Cluster 1 compared to Cluster 2 ($p = 0.003$). However, when controlled for gender, this difference lost its statistical significance ($p = 0.0475$, did not survive Holm-Bonferroni correction for multiple comparisons).

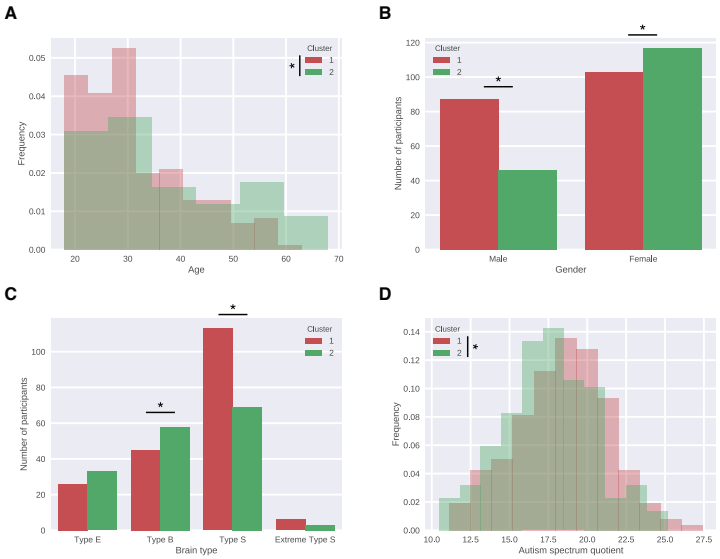

Figure 3.3: Significant differences between the two clusters. ***A***: Distribution of age in the two clusters. ***B***: Number of participants with each gender in the two clusters. ***C***: Number of participants with each cognitive brain type in the two clusters. ***D***: Distribution of AQ in the two clusters.

Next, a logistic regression was performed to confirm the effects of age, gender, brain type and AQ on the assignment of participants to each cluster (Table 3.4). The logistic regression model was statistically significant, $\chi^2(6) = 36.55, p < .001$. The model explained 13% (Nagelkerke R^2) of the variance in the cluster assignment and correctly classified 64% of cases, indicating that there are other unaccounted factors influencing the cluster assignment. Age

($\beta = 0.04$, $p < 0.001$) and gender ($\beta = 0.57$, $p = 0.02$) added significantly to the model, whereas brain types ($p = 0.57$) and AQ ($p = 0.99$) did not have a significant effect. According to the model, males were 1.77 times more likely to be assigned to the Cluster 1 (high complexity preference) than females. Furthermore, increasing age was associated with an increased likelihood of assignment to Cluster 2 (low complexity preference).

Table 3.4: Results of logistic regression with age, gender, brain type and AQ variables. Abbreviations: SE = standard error, DF = degrees of freedom, CI = 95% confidence interval for odds ratio ($\exp\beta$).

	β	SE	DF	p-value	odds ratio	Lower CI	Upper CI
Age	0.04	0.01	1	0.000	1.04	1.02	1.06
Gender(1)	0.57	0.25	1	0.022	1.77	1.08	2.89
AQ	0.00	0.08	1	0.998	0.99	0.85	1.17
Brain type	-	-	3	0.565	0.57	-	-
Brain type(1)	-0.20	0.86	1	0.817	0.82	0.15	4.39
Brain type(2)	0.50	0.59	1	0.400	1.65	0.51	5.27
Brain type(3)	0.52	0.37	1	0.160	1.68	0.81	3.48
Constant	-2.12	1.64	1	0.198	0.12	-	-

3.4 Discussion

In this study, we showed that the previously observed duality of complexity preferences (Güçlütürk et al., 2016b) was not only limited to the visual domain and highly controlled stimuli, but also was evident in the auditory domain with complex ecologically valid music stimuli. We demonstrated the importance of accounting for individual differences by revealing opposite complexity preference patterns for two groups of participants in a sample of over 300 participants. Furthermore, we characterized the identified groups with different complexity preferences with post-hoc analyses and demonstrated that the distribution of age, gender, cognitive brain type and autism spectrum quotient differed between the groups that preferred complex or simple songs. Similar to previous reports regarding music preferences (North, 2010),

the results of our analyses suggest that demographic measures were the most important variables predicting complexity preferences. Specifically, we found that younger people were more likely to prefer complex songs whereas for older people it was the opposite. Furthermore, males were more likely to be assigned to the group with high complexity preference, whereas females were more likely to prefer simpler songs.

The result that grouping the participants increased the ICC as well as improving the fit of the linear mixed models demonstrates the necessity of using simple yet effective methods like clustering for evaluating the effects of modulating factors of liking. Our results show that separating this relatively large sample of participants into two groups reveals the best grouping, and on average these two groups have opposite complexity preferences. However, it is important to be aware of the further variability within these two groups that may be driven by factors other than complexity. Although the two opposite liking vs. complexity functions are evident even upon visual inspection in Figure 3.2 (as well as quantitatively in Tables 3.1 and 3.2), the linear relationships between liking and complexity in the clusters are still noisy. We believe that such variability in preferences is expected as the stimulus songs varied in the many dimensions of the MUSIC model (Rentfrow et al., 2011) and were not controlled in many aspects. Besides introducing some level of noise/variability to the results, another consequence of using such an uncontrolled (and ecologically valid) stimulus set is that it allows making general conclusions spanning the large extent of the music domain. While the current results provide evidence for a general relation between complexity and liking in music, investigating within genre preferences with similar methods would be an interesting next step in the study of complexity preferences for music.

With respect to the characterization of participants with simple and complex music preferences, the observed age differences between the two clusters of participants is an interesting but not

an unexpected result. Very recently, Pugach, Leder and Graham (2017) showed that visual aesthetic preferences are not stable across the lifespan. Our results conform to this finding and suggest that older participants were more likely to prefer simpler songs whereas the younger participants were more likely to prefer more complex songs. In the visual domain, such a relationship has been suggested earlier (Alpaugh & Birren, 1977; Crosson & Robertson-Tchabo, 1983; Güçlütürk et al., 2016b; Munsinger et al., 1964). In the auditory domain, an important role of age in music preferences has been established (Bonneville-Roussy, Rentfrow, Xu & Potter, 2013; Bonneville-Roussy, Stillwell, Kosinski & Rust, 2017), however these investigations were more focused on music genres rather than the complexity dimension of music. Although it is difficult to disentangle the impact of complexity on the perception and categorization of musical genres, it is important to also consider possible effects of age-related differences in genre preferences. There are only a few studies that investigated gender differences in musical complexity preferences (Marin & Leder, 2013), Previously, Marin and Leder (2013) showed that for females, pleasantness and complexity of classical music excerpts were negatively correlated, whereas for males, they were positively correlated. The results of the current study are in line with these results, and thus emphasizes gender as an important factor. Future studies should further investigate its role in complexity preferences.

Our initial analysis showed that there were small but significant differences between the two clusters in terms of their average AQ, such that people that were assigned to the high complexity preference cluster had a significantly higher average AQ. However, the results of the logistic regression analysis suggest that this difference was likely driven by the gender differences. Nevertheless, it would be interesting to investigate a possible link between superior or abnormal sensory processing as in ASC (Dakin & Frith, 2005; Haesen et al., 2011; Robertson & Simmons, 2012) and preferences to better explain the current results in connection to

recent findings regarding strong links between visual sensitivity and preferences (Spehar et al., 2015).

We believe that the study of aesthetics can benefit from simple yet effective approaches that we illustrated in the current study. Pooling preference data across participants may easily obscure relevant differential tendencies between groups of participants. Diving into the nature of these differences is likely to reveal new insights in the underlying mechanisms and interindividual differences driving these data.

On top of the above mentioned future research directions, the approach and the results of the current study generate several new research questions as listed below.

- Is the duality of complexity preferences limited to simple visual patterns and music of various genres or can it be observed with other sets of stimuli, for example within a specific genre of songs, or a varied set of paintings? To what extent do these results generalize?

- Does the preference in a modality also depend on sensory sensitivity, and if so, what type of sensitivity could potentially account for these differences?

- Does the complexity preference in one modality also persist in the other modality, i.e. if a person likes complex music, would they also like complex visual art? If this is the case, this may suggest a different supramodal mechanism and an explanation other than sensory sensitivity. Such results would further necessitate moving towards neuroaesthetics theories encompassing different sensory modalities (Brattico, Brattico & Vuust, 2017; Marin, 2015).

Supplementary Materials

Table 3.5: List of songs in the order of increasing complexity

	Artist	Song	Genre
q5	Human Signals	Birth	soft rock
q4	Language Room	She Walks	soft rock
q25	Ali Handal	Sweet Scene	soft rock
q1	Lisa McCormick	Let's Love	adult contemporary
q23	Bob Delevante	Penny Black	new country
q11	Ljova	Seltzer, do I drink too much	avant-garde classical
q24	Curtis	Carrots and Grapes	rock-n-roll
q3	Kush	Sweet 5	electronica
q22	Carey Sims	Praying for Time	mainstream country
q2	Leo The Lionheart	050107 electro	electronica
q15	Moh Alileche	North Africa's Destiny	world beat
q6	AB+	Recess	electronica
q10	Ciph	Brooklyn Swagger	rap
q13	Lisa McCormick	Fernando Esta Feliz	latin
q21	Anglea Motter	Mama I'm Afraid To	bluegrass
q14	Daniel Nahmod	I Was Wrong	traditional jazz
q12	DNA	La Wally	classical
q7	The Cruxshadows	Go Away	europop
q16	Exit 303	Falling Down 2	classic rock
q17	Cougars	Dick Dater	classic rock
q19	The Stand In	Frequency of a Heartbeat	punk
q18	Five Finger Death Punch	White Knuckles	heavy metal
q20	Straight Outta Junior High	Over now	punk

4

Representations of Naturalistic Stimulus Complexity in Early and Associative Visual and Auditory Cortices

This chapter is based on Güçlütürk, Y., Güçlü, U., van Gerven, M., and van Lier, R. (2018). Representations of naturalistic stimulus complexity in early and associative visual and auditory cortices. Scientific Reports, 8:3439. https://doi.org/10.1038/s41598-018-21636-y

4.1 Introduction

Early research in perception identified stimulus complexity as one of the most important stimulus properties that affect task performance(Donderi, 2006). In the visual domain, stimulus complexity has been shown to influence time perception(Schiffman & Bobko, 1974), speed and accuracy of shape recognition(Kayaert & Wagemans, 2009), amodal completion(van Lier, 1999; de Wit, Bauer, Oostenveld, Fries & van Lier, 2006) the reaction time of search, discrimination and recognition tasks(Fitts et al., 1956; Anderson & Leonard, 1958; Mavrides & Brown, 1969; Donderi & McFadden, 2005; Koning & Lier, 2005), as well as affecting memory(Simon, 1972; Chai, 2010) and perceptual learning (Ahissar & Hochstein, 1997; Hazenberg, Jongsma, Koning & van Lier, 2014) performance. Furthermore, complexity relates to interestingness, pleasantness, liking and similar subjective aesthetic evaluations of art(Berlyne, 1971), natural images of scenes(Stamps, 2004) and architecture(Herzog & Shier, 2000). It has been much studied in the domain of art perception and empirical aesthetics(**Gucluturk2016**; Nadal et al., 2010; Marin & Leder, 2013; Braun et al., 2013; Muth et al., 2015) as well as environmental psychology(Stamps, 2004; Tveit, Ode & Fry, 2006) for its role in preferences. Recently, its effect on neural decoding of visual stimuli from brain activity has also been investigated(Güçlütürk et al., 2017). Similarly in the auditory domain, complexity plays an important role in stimulus perception with its modulatory effects on attention, arousal and memory (Potter & Choi, 2006). Furthermore, as is the case with its visual counterpart, music complexity highly influences whether a song will be liked or not(Orr, 2005; Marin & Leder, 2013, 2016; Heyduk, 1975; North & Hargreaves, 1995).

Borrowing ideas from information theory(Shannon, 1948), complexity of a stimulus is commonly thought as the amount of information contained in it(Shmulevich & Povel, 2000). However,

moving away from information theory, when the perceptual limitations of humans are embedded into its definition, complexity becomes a subjective and multi-faceted concept(Berlyne et al., 1968; Donderi, 2006; Nadal et al., 2010). In the case of visual complexity, the dimensions making up the complexity of a stimulus include properties such as regularity, number of elements, and diversity(Berlyne, 1971). Similarly, according to structural information theory, complexity (or simplicity) is defined in terms of the regularity of patterns, which depends on iteration, symmetry and alternation properties(Leeuwenberg, 1969; van der Helm, 2014; van Lier, 2001). However, for natural images such as photographs of scenes, objects or works of art, it is difficult to correctly determine such properties. In these cases, computational measures of complexity become useful and can be evaluated automatically instead of using the aforementioned elements to define an intractable complexity measure. Specifically, measures related to information content such as Kolmogorov complexity estimated as compressed file size of images(Donderi & McFadden, 2005; Marin & Leder, 2013; Machado et al., 2015; Forsythe et al., 2011), self-similarity and fractal dimension(Cutting & Garvin, 1987; Corbit & Garbary, 1995; Spehar et al., 2003; Taylor et al., 2005), and Pyramid of Histograms of Orientation Gradients (PHOG) derived measures(Redies et al., 2012; Braun et al., 2013) have been frequently used to automatically obtain an estimate of subjective complexity of images. In the context of music, complexity is a highly subjective term that encompasses several properties(Shmulevich & Povel, 2000), such as tempo, predictability, variety of instruments, harmony and rhythm(Streich, 2007). However, recent models of music complexity suggest that the best predictors of subjective complexity of a song are the event density and Kolmogorov complexity estimated as the compressed file size of the song(Marin & Leder, 2013).

Despite its importance, functional neuroimaging studies investigating the neural correlates of stimulus complexity have been limited, especially for naturalistic stimuli. In the visual domain, electro-

physiological investigations showed that the amplitude of the N2 event-related potential and late positive potential increased with increased complexity of randomly generated polygons(Shigeto et al., 2011). One of the first functional brain imaging studies that measured cerebral metabolic rate for glucose using positron emission tomography (PET) determined that as stimulus complexity increased (from white light to an alternating checkerboard pattern) the glucose metabolic rate also increased in both the primary and associative visual cortices, but faster in the associative visual cortex(Phelps et al., 1981). Similarly, using PET, it was shown that cerebral blood flow velocity in the occipital cortex increased as visual stimuli from different categories with increasing complexity (diffuse light, checkerboard pattern and movie) is presented to subjects(Sitzer et al., 1992). In the case of auditory stimulus complexity, a meta-analysis investigated where the different categories of sounds such as pure tones, noise, music and vocal sounds caused activations across the human brain(**Samson2011**). In a different study, it was shown that brain connectivity patterns of people listening to their favorite songs were consistent with each other, despite the differences between their favorite songs in terms of complexity(Wilkins et al., 2014).

The current study aims to provide a comprehensive account of representations of naturalistic sensory stimulus complexity in sensory cortices by establishing a direct, predictive relationship between objective stimulus complexity measures and stimulus-evoked brain activations as well as characterizing the properties of this relationship under the framework of neural encoding and decoding in functional magnetic resonance imaging (fMRI) (Naselaris, Kay, Nishimoto & Gallant, 2011). That is, we develop several neural encoding and decoding models, which embody specific hypotheses about certain stimulus features modulating stimulus-evoked brain activations, to test alternative hypotheses about what, if any, stimulus complexity measures are represented in different brain regions as well as analyzing these models to characterize the properties of these representations. To this end,

we present the results of two different fMRI experiments; one with naturalistic image stimuli to characterize the representations of visual complexity and another with music stimuli to determine the those of auditory/music complexity in the brain. Our results reveal four core findings regarding the processing of stimulus complexity in the sensory cortices: i) Stimulus complexity was shown to modulate visual and auditory cortices. ii) A quantification of the complexity sensitivity of individual regions of interest (ROIs) demonstrated a change of sensitivity (from fine grained to coarser) in a gradient from lower to higher areas. iii) It was shown that parahippocampal place area (PPA) has distributed representations of complexity comparable to or better than the ROIs in the early visual cortex, supporting the notion that global scene properties such as complexity plays an important role in scene processing. iv) It was shown that regions of the auditory cortex, which represent syntactic language complexity such as posterior regions of superior temporal gyrus (STG) and superior temporal sulcus (STS), also represent music complexity.

4.2 Materials and Methods

Ethics Statements

Experiment 1 was approved by the Ethics Committee of ATR and the subjects provided written informed consent. Experiment 2 was approved by the local ethics committee (CMO Regio Arnhem-Nijmegen) and the subjects provided written informed consent in accordance with the Declaration of Helsinki. All experiments were performed in accordance with the relevant guidelines and regulations.

Experiment 1

Dataset

In the first experiment, we used a preexisting dataset(**Horikawa2017**). Here, we report only the most pertinent details. Additional details can be found in the original publication. The dataset comprises visual stimuli (photographs) and fMRI data of five healthy adult subjects (23-38 year old one female and four male subjects with normal or corrected-to-normal vision).

Design The dataset comes in two parts [The dataset contains also an imagery set, which is not considered here.]:

i) A training set that was collected in 25 unique runs over multiple sessions, which consist of 50 unique and 5 repeated trials per run. In one training run, 51 different images were presented. Fifty of these images were repeated only once within the run. One of these images were repeated five times within the run. Each of the 25 training runs used a different set of images.

ii) A test set that was collected in 35 repeated runs over multiple sessions, which consist of 50 unique and 5 repeated trials per run. Like the training set, in one test run, 51 different images were presented. Fifty of these images were repeated only once within the run. One of these images were repeated five times within the run. Unlike the training set, each of the 35 test runs used the same set of images.

The repeated images were used to facilitate a one-back task and were not included in the final data set. After the exclusion of the repeated images, the training set ended up with 1250 different images repeated once each, while the test set ended up with 50 different images repeated 35 times each.

Each image was presented at 12 ° × 12 ° and 5 Hz for 9 s, which was followed by the one-back repetition detection task. Subjects fixated a central point throughout each run.

Visual stimuli The stimuli were drawn from the subset of the original ImageNet dataset(**JiaDeng2009**) that was used in the Large Scale Visual Recognition Challenge 2011(Russakovsky et al., 2015) (ILSVRC2011). The subset comprises 1350000 photographs, each of which belongs to one out of 1000 categories. The training set contains 1200 stimuli that belong to 150 representative categories (eight photographs per category). The test set contains 50 stimuli that belong to 50 representative categories (one photograph per category). The categories are mutually exclusive. A list of these 200 categories are available as Supplementary Information. Example categories include bathtub, chimpanzee, fire truck, human being, hot-air balloon, watermelon, etc. Each stimulus was center cropped to the largest square possible and resized to 500 px × 500 px with antialiasing and bicubic interpolation.

MRI data MRI data were acquired with a Siemens 3 T MAGNETOM Trio scanner. Anatomical scans were collected with T1-weighted MP RAGE and T2-weighted turbo spin echo pulse sequences. Functional scans were collected with a T2*-weighted gradient echo echo planar imaging pulse sequence (voxel size: 3 mm^3; slices: 50 for localizer and task scans, and 30 for retinotopy scans; distance factor: 0 %; FoV read: 192 mm; TR: 3000 ms for localizer and task scans, and 2000 ms for retinotopy scans; TE: 30 ms; flip angle: 80 degrees; multi-slice mode: interleaved).

The fMRI data were preprocessed as follows: The functional scans were realigned to one another and coregistered to the anatomical scans. The realigned and coregistered functional scans in each run were linearly detrended and standardized along the time axis. The realigned functional scans in each trial were shifted by 3 s

(one TR), cropped to the first 9 s (3 TRs) and averaged along the time axis. This resulted in one time point per trial per voxel.

Seven ROIs in the lower visual cortex and the higher visual cortex were defined based on retinotopic mapping and functional localizers. Table 4.1 shows the details of these ROIs.

Table 4.1: ROIs considered in the first (visual) experiment. 1: (Engel et al., 1994); 2: (Sereno et al., 1995); 3: (Kourtzi & Kanwisher, 2000); 4: (Kanwisher, McDermott & Chun, 1997); 5: (Epstein & Kanwisher, 1998).

Area	Also known as	Region	Reference
V1 (Primary Visual Cortex)	17, hOC1, OC, BA17	Lower Visual Cortex	1,2
V2 (Second Visual Area)	18, hOC2, OB, BA18	Lower Visual Cortex	1,2
V3 (Third Visual Area)	V3d, V3v, VP, hOC3d, hOC3v	Lower Visual Cortex	1,2
V4 (Fourth Visual Area)	V4d, V4v, hV4, hOC4v, hOC4lp, LO1	-	1,2
LOC (Lateral Occipital Complex)	LO1, LO2, hOC4la	Higher Visual Cortex	3
FFA (Fusiform Face Area)	FFC, FG2	Higher Visual Cortex	4
PPA (Parahippocampal Place Area)	-	Higher Visual Cortex	5

Visual Complexity Measures

The complexity of the visual stimuli was parameterized with three different measures. These computational visual complexity measures are objective measures of complexity, which are estimates of subjective complexity. These exact measures have been shown to reflect the subjective complexity levels of images in earlier behavioral studies(**Gucluturk2016**; Redies et al., 2012; Braun et al., 2013).

- **Mean maximum magnitude gradient (Gradient):** This measure was based on the 'Complexity' measure as described in a previous study(Redies et al., 2012). It measures the maximum rate of change in the Lab color space channels of an image. It was computed as follows(Redies et al., 2012):

$$\text{mean}_{x,y} \max_{c} ||\nabla \text{S}_c(x, y)|| \tag{4.1}$$

where $S_c(x, y)$ is the pixel (x, y) in the channel c of the stimulus in the Lab color space such that $c \in \{L, a, b\}$, $x \in \{1, \ldots, 256\}$ and $y \in \{1, \ldots, 256\}$.

- **Portable Network Graphics (PNG):** The PNG-based complexity measure of a stimulus was computed as the compressed (lossless) file size (bytes) of the stimulus in the RGB color space. This measure can be thought of as a surrogate for the Kolmogorov complexity of the image, which is defined as the length of the shortest computer program or algorithm that can be used to represent the object (Solomonoff, 1986). It essentially gives an indication of the information content of the image. PNG encoding was carried out using *Pillow 4.1.1* (https://python-pillow.org/) with default parameter settings.

- **Self-similarity:** Self-similarity is a measure of how much the whole of an object resembles its parts. To estimate this for each stimulus image, we compared histograms of oriented gradients (HOGs) of the whole image at the ground level and subregions of the image at the third level(Redies et al., 2012). Specifically, the self similarity-based complexity measure of a stimulus was computed as follows(Redies et al., 2012):

$$\text{median}_f\left(\sum_{x=1}^{8}\sum_{y=1}^{8}\min\left(\text{PHOG}^{(2)}_{\lceil x/2\rceil,\lceil y/2\rceil}(f), \text{PHOG}^{(3)}_{x,y}(f)\right)\right) \tag{4.2}$$

where $\text{PHOG}^{(l)}_{x,y}(f)$ is the feature f in the subregion (x, y) and the level l of the pyramid histogram of oriented gradients (Dalal & Triggs, 2005; Bosch et al., 2007) of the stimulus in the Lab color space such that $f \in \{1, ..., 16\}$ (16

equally spaced orientations between $-\pi$ radians and π radians).

Note that all visual stimuli were resized to 256 px $\times$ 256 px with Lanczos interpolation and converted to the RGB color space (if required) prior to the computation of the visual complexity measures.

Experiment 2

For the second experiment, we used a new dataset, which comprises auditory stimuli (music) and fMRI data of eight healthy adult subjects (24-38 year old four female and four male subjects with normal hearing).

Design

Subjects participated in two sessions: one for a training set and another one for a test set. The training set was collected in eight unique runs, each of which consisted of 16 unique trials. The test set was collected in eight repeated runs, each of which consisted of 16 unique trials.

In each trial, a stimulus was presented for 29 s with in-ear headphones, followed by a self-paced complexity preference rating task. Subjects fixated a central point throughout each trial. The task was used to keep the subjects engaged, and the ratings were excluded from the analyses.

Prior to entering the scanner, subjects listened to three example stimuli: a very loud one, a normal one and a very quite one. After subjects entered the scanner and before the first run, the second example stimulus was presented for 29 s with in-ear headphones (while fMRI data were acquired, which were later discarded).

During this period, subjects adjusted the volume to a comfortable level.

Auditory stimuli

We used 144 auditory stimuli that were systematically drawn from the MagTag5k Autotagging Dataset(Marques, Domingues, Langlois & Gouyon, 2011) – a preprocessed version of the original MagnaTagATune Dataset(Law, West, Mandel, Bay & Downie, 2009), which solves some of the problems in the original such as duplication and synonymy. The preprocessed dataset contains 5259 29-second long, 16000 Hz song excerpts and 136 binary tags per excerpt.

In order to create a stimulus dataset that spanned a large musical spectrum based on their associated tags, We used hierarchical (agglomerative) clustering with correlation distance (1 - Pearson correlation coefficient) and complete linkage to cluster all song excerpts to 16 clusters based on the binary tags. Among all excerpts in each cluster, the one with the highest within-cluster similarity was assigned to the test set. This resulted in a test set of 16 stimuli (= 16 clusters $\times$ 1 excerpt). Among remaining excerpts in each cluster, eight with the highest within-cluster similarity were assigned to the training set. This resulted in a training set of 128 stimuli (= 16 clusters $\times$ 8 excerpts). A list of all 59 tags that were associated to the 144 stimuli are available as Supplementary Information. Example tags include ambient, electro, instrumental, female vocal, piano, strange, etc.

MRI data

MRI data were acquired with Siemens 3 T MAGNETOM Prisma^{fit} scanner and Siemens 32-Channel Head Coil. Anatomical scans

were collected with a T1-weighted MP RAGE pulse sequence (voxel size: 1 mm^3; slabs: 1; distance factor: 50 %; orientation: sagittal; FoV read: 256 mm; slice thickness: 1 mm; TR: 2300 ms; TE: 3.03 ms; PAT mode: GRAPPA; accel. factor PE: 2; flip angle: 8 degrees; multi-slice mode: single shot). Functional scans were collected with a T2*-weighted gradient echo echo planar imaging pulse sequence (voxel size: 2 mm^3; slices: 64; distance factor: 0 %; orientation: transversal; FoV read: 210 mm; slice thickness: 2.4 mm; TR: 735 ms; TE: 39 ms; multi-band accel. factor: 8; flip angle: 75 degrees; multi-slice mode: interleaved).

The fMRI data were preprocessed as follows: The functional scans were realigned to the first functional scan and the mean functional scan, respectively. Realigned functional scans were slice time corrected. The realigned and slice time corrected functional scans in each run were linearly detrended and standardized along the time axis. The realigned and slice time corrected functional scans in each trial were shifted by 2.94 s (four TRs), cropped to 29.4 s (40 TRs; approximately one stimulus) and averaged with a window size of 8.82 s (12 TRs) and a hop size of 0.735 s (one TR) along the time axis. This resulted in 28 time points per trial.

Thirteen ROIs in the early auditory cortex and the auditory association cortex were defined based on the HCP MMP 1.0 parcellation(Glasser et al., 2016) after projecting it to the native volumetric space via HCP MMP 1.0 parcellation $\rightarrow$ fsaverage surface space $\rightarrow$ native surface space $\rightarrow$ native (anatomical) volumetric space $\rightarrow$ native (functional) volumetric space. Table 4.2 shows the details of these ROIs.

Auditory Complexity Measures

The complexity of the auditory stimuli was parameterized with three different measures. These computational auditory complexity measures are objective measures of complexity, which are

Table 4.2: ROIs considered in the second (auditory) experiment. 1: (Glasser et al., 2016); 2: (Glasser & Essen, 2011); 3: (Moerel, Martino & Formisano, 2014); 4: (Triarhou, 2007b); 5: (Triarhou, 2007a); 6: (Pandya & Sanides, 1973); 7: (Kurth et al., 2009); 8: (Morosan, Schleicher, Amunts & Zilles, 2005); 9: (Morosan et al., 2001).

Area	Also known as	Region	Reference
A1 (Primary Auditory Cortex)	Core, R1, TC, TE1.0, TE1.1, 41	Early Auditory Cortex	1,2,3,4,5
LBelt (Lateral Belt Complex)	Belt, TB	Early Auditory Cortex	1,3,4,5
MBelt (Medial Belt Complex)	Belt, TB	Early Auditory Cortex	1,3,4,5
PBelt (ParaBelt Complex)	ParaBelt, TA1	Early Auditory Cortex	1,3,4,5
RI (RetroInsular Cortex)	reI, reIt, RetroInsular, Belt, TD	Early Auditory Cortex	1,2,6,7,4,5
A4 (Auditory 4 Complex)	TE3	Auditory Association Cortex	1,8
A5 (Auditory 5 Complex)		Auditory Association Cortex	1
STSdp (Area STSd posterior)		Auditory Association Cortex	1
STSda (Area STSd anterior)		Auditory Association Cortex	1
STSvp (Area STSv posterior)		Auditory Association Cortex	1
STSva (Area STSv anterior)		Auditory Association Cortex	1
STGa (Area STGa)		Auditory Association Cortex	1
TA2 (Area TA2)	TE1.2	Auditory Association Cortex	1,4,5,9

estimates of subjective complexity. These exact measures have been shown to reflect the subjective complexity levels of songs in an earlier comprehensive behavioral study analyzing the relationship between various subjective and objective stimulus complexity measures(Marin & Leder, 2013).

- **Free Lossless Audio Codec file size (FLAC):** The compressed file size (in bytes) of the stimuli with a lossless audio codec. This measure can be thought of as the Kolmogorov complexity of the stimuli, similar to the PNG measure in the case of the visual stimuli. FLAC encoding was carried out using *ffmpeg* (https://ffmpeg.org/) with 16000 samples per second and 16 bits per sample.

- **Ogg Vorbis file size (Ogg):** The compressed file size (in bytes) of the stimuli with a lossy audio codec. This is another type of Kolmogorov complexity estimate. While FLAC utilizes a lossless compression method, Ogg Vorbis uses an audio coding format which results in lossy compression. Ogg encoding was carried out using *ffmpeg* (https://ffmpeg.org/) with 16000 samples per second and 75 quality.

- **Event density:** The mean frequency (in hertz) of simultaneous harmonic, melodic and rhythmic events in the stimuli. As an example, a song with a fast rhythm would have a higher event density compared to a song with a slow rhythm. Event density was estimated with *MIRtoolbox* (https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox).

All measures were extracted from each 29-second long, 16000 Hz stimuli using a sliding window analysis with a window size of 8.82 s (12 TRs) and a hop size of 0.735 s (one TR). This resulted in 28 time points per stimulus per measure, allowing us to analyze the complexity processing in a dynamic fashion (Muth et al., 2015).

Decoding Analysis

In the decoding analysis, we used ridge regression to predict complexity measures from stimulus-evoked responses of voxels in ROIs. Let $x \in \mathbb{R}$ and $\mathbf{y} \in \mathbb{R}^q$ be a pair of a complexity measure and stimulus-evoked responses of q voxels in an ROI. We are interested in predicting x as a linear function of $\mathbf{y}$:

$$x = \boldsymbol{\beta}^\top \mathbf{y} \tag{4.3}$$

where $\boldsymbol{\beta}$ is regression coefficients. Without loss of generality, we assume that x and $\mathbf{y}$ have zero mean and unit variance. We minimize the L^2 penalized least squares loss function to estimate $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{\beta}} \left[\frac{1}{n} \sum_{i=1}^{n} \left(x^{(i)} - \boldsymbol{\beta}^\top \mathbf{y}^{(i)} \right) + \lambda ||\boldsymbol{\beta}||_2^2 \right] = \left(\mathbf{Y}^\top \mathbf{Y} + \lambda \mathbf{I}_q \right)^{-1} \mathbf{Y}^\top \mathbf{x} \tag{4.4}$$

where $\lambda \geq 0$ is a regularization coefficient and $\{\mathbf{x}, \mathbf{Y}\}$ with $\mathbf{x} = \left(x^{(1)}, \ldots, x^{(n)}\right)^\top$ and $\mathbf{Y} = \left(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}\right)^\top$ is a training set consisting of n x-$\mathbf{y}$ pairs.

We used grid search to optimize λ as follows(Güçlü & van Gerven, 2014): First, 100 linearly spaced values between 0.1 and $\min(n, q) - 0.1$ were used to specify a set of effective degrees of freedom of the ridge regression fit. Then, Newton's method was used to solve each value for λ. Finally, leave-one-out cross-validation on the training set was used to choose λ.

We estimated a separate decoding model for each complexity measure and all voxels in each ROI. We validated the decoding models on a held-out test set, which was at no point used for model estimation. We used the Pearson correlation coefficient between the ground truths and the predictions on the test set (r) as the validation metric.

Encoding Analysis

In the encoding analysis, we use linear regression to predict stimulus-evoked responses of voxels in ROIs from complexity measures. Let $x \in \mathbb{R}$ and $y \in \mathbb{R}$ be a pair of a complexity measure and a stimulus-evoked response of a voxel in an ROI. We are interested in predicting y as a linear function of x:

$$y = \beta^\top x \tag{4.5}$$

where β is a regression coefficient. Without loss of generality, we assume that x and y have zero mean and unit variance. We minimize the least squares loss function to estimate β:

$$\beta = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} (y^{(i)} - \beta^{\top} x^{(i)}) \tag{4.6}$$

$$\beta = (\mathbf{x}^{\top}\mathbf{x})^{-1}\mathbf{x}^{\top}\mathbf{y} \tag{4.7}$$

where $\mathbf{x} = (x^{(1)}, \ldots, x^{(n)})^{\top}$ and $\mathbf{y} = (y^{(1)}, \ldots, y^{(n)})^{\top}$ is a training set with n x-y pairs.

We estimate a separate encoding model for each complexity measure and each voxel in each ROI. We validate the encoding models on a held-out test set, which is at no point used for model estimation. We use the Pearson correlation coefficient between the ground-truths and the predictions on the test set (r) as the validation metric.

Statistical Analysis

We use permutation testing to test the null hypothesis that r of a model is not different than the chance level as follows: First, the order of the ground-truths in the training and test sets are randomly shuffled. Then, a new model is estimated on the new training set and validated on the new test set. These steps are repeated 1000 times. The chance level is taken to be mean r of the new models. The p-value is taken to be the fraction of the new models whose r is greater than or equal to r of the old model. The null hypothesis is rejected if the p-value is less than or equal to 0.05.

Hyperalignment

We perform the analyses on mean hyperaligned fMRI data (Haxby et al., 2011) (SH), which are obtained via the following iterat-

ive process: Before the first iteration, the training set fMRI data of the subject who has the most number of voxels are assigned to a common representational space. At each iteration, i) the training set fMRI data of each subject are transformed to the common representational space with Procrustes transformation, ii) averaged along the subjects and iii) reassigned to the common representational space. After the final iteration, i) the training and test set fMRI data of each subject is transformed to the common representational space with Procrustes transformation and ii) averaged along the subjects. Note that the fMRI data from different experiments and/or ROIs are hyperaligned separately (Tables 4.3 and 4.4)..

In other words, this process minimizes the Procrustes distance between fMRI data of different subjects with geometric transformations (rotation, translation and/or uniform scaling), which change the placement and the size but preserve the shape of the fMRI data of the different subjects. For example, consider the following toy three-trial fMRI data of two subjects who have two voxels each: {(1, 2), (3, 4), (5, 6)} and {(-24, 16) (-45, 35), (-66, 54)}. The former can be aligned to the latter with these transformations while keeping the shape of the former (line) intact as follows:

Uniform scaling by 10: **{(1, 2), (3, 4), (5, 6)}** $\rightarrow$ {(10, 20), (30, 40), (50, 60)}
Translation by 5: {(10, 20), (30, 40), (50, 60)} $\rightarrow$ {(15, 25), (35, 45), (55, 65)}
Rotation by $\pi/2$ radians: {(15, 25), (35, 45), (55, 65)} $\rightarrow$ **{(-25, 15) (-45, 35), (-65, 55)}**

which results in a Procrustes distance (root sum square) of 2.

Table 4.3: Number of voxels in ROIs of the subjects from Experiment 1. SH denotes mean hyperaligned fMRI data.

	V1	V2	V3	V4	LOC	FFA	PPA
S1	**1004**	1018	759	740	540	568	356
S2	757	944	810	544	834	435	316
S3	872	**1031**	861	754	**996**	928	496
S4	719	855	**929**	704	668	725	398
S5	659	891	907	**860**	566	**929**	**550**
SH	**1004**	**1031**	**929**	**860**	**996**	**929**	**550**

Table 4.4: Number of voxels in ROIs of the subjects from Experiment 2. SH denotes mean hyperaligned fMRI data.

	A1	LBelt	MBelt	PBelt	RI	A4	A5	STSdp	STSda	STSvp	STSva	STGa	TA2
S1	40	71	73	119	87	241	269	177	**288**	201	191	183	117
S2	39	67	71	87	92	162	172	131	142	153	138	138	112
S3	**81**	90	**128**	163	**111**	320	285	**206**	197	**274**	**200**	154	**165**
S4	70	**103**	109	**173**	102	**322**	**288**	197	220	227	199	143	147
S5	54	95	89	142	93	244	254	151	198	189	180	146	126
S6	33	59	87	104	71	221	173	144	131	187	118	132	128
S7	51	91	88	129	82	258	249	139	207	251	164	**183**	99
S8	62	83	98	132	89	254	281	163	205	181	156	139	140
SH	**81**	**103**	**128**	**173**	**111**	**322**	**288**	**206**	**288**	**274**	**200**	**183**	**165**

Data Availability

The dataset analysed during the current study (Experiment 1) is available in the ATR brainliner repository, http://brainliner.jp/data/brainliner/Generic_Object_Decoding. The dataset generated and analysed during the current study (Experiment 2) is available from the corresponding author on reasonable request.

4.3 Results

Experiment 1

We first examined the relationship between the measures of visual complexity by calculating bivariate Pearson correlation coefficients between each measure (Table 4.5). All three measures were significantly correlated with each other ($p < 0.05$). Highest correlation was between gradient and PNG ($r = 0.86$), whereas the lowest correlation was observed between gradient and self-similarity ($r = 0.48$).

Table 4.5: Experiment 1 - Bivariate correlations (r) between visual complexity measures.

	Gradient	**PNG**	**Self-similarity**
Gradient	1.00	0.86	0.48
PNG	0.86	1.00	0.65
Self-similarity	0.48	0.65	1.00

Next, we performed the decoding analysis on the hyperaligned data of the five subjects. For each visual complexity measure, we predicted the value of the complexity measure from the stimulus-evoked responses of voxels in the ROIs from the visual cortex (Table 4.2) using ridge regression. Note that this is a multivariate analysis, such that in order to predict the values of a complexity measure of stimuli, all voxel responses in a ROI are used at the same time. Figure 4.1 shows the visual complexity decoding results. Remarkably, from all ROIs in the visual cortex, it was possible to predict all three visual complexity measures significantly above chance level ($p \ll 0.001$) with a maximum correlation between the predicted values and actual complexity measure values of $r = 0.75$ for PNG measure in PPA and a minimum correlation of $r = 0.40$ for self-similarity measure in FFA. Among all three complexity measures, PNG had the highest ($r = 0.67$) and self-similarity had the lowest ($r = 0.56$) average correlation over the ROIs. For all three measures, LOC and FFA regions had the lowest decoding performance, and among the ROIs in the higher visual

cortex, all complexity measures could be predicted with highest accuracy from PPA (r = 0.65, 0.75, 0.62 for gradient, PNG and self-similarity measures, respectively).

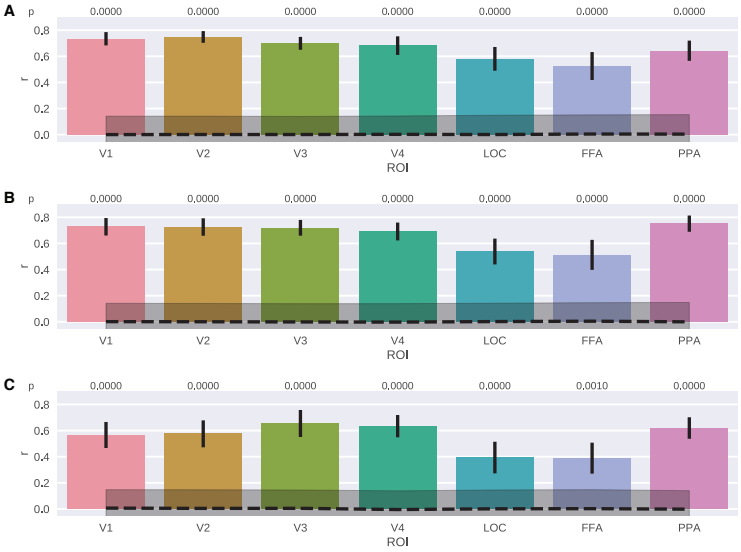

Figure 4.1: Experiment 1 - Results of decoding visual complexity measures from the ROIs in the visual cortex. ***A***: Gradient. ***B***: PNG. ***C***: Self-similarity. Bars and error bars show decoding performance and ± 1 SEM, respectively. Dashed line and shaded region show chance level and ± 1 SEM around r = 0, respectively. SEM is computed with bootstrapping (1000 iterations).

Then, we evaluated how well the stimulus-evoked responses in ROIs could be predicted from the visual complexity measures using linear regression (Figure 4.2). Note that unlike the decoding analyses which were multivariate, encoding analyses are univariate, such that we estimate a separate encoding model for each complexity measure and each voxel in each ROI. For all complexity measures, we observed that the encoding performance decreased along the visual hierarchy such that the percentage of voxels whose responses were predicted significantly above chance ($p < 0.05$) was highest in V1 (65%, 62% and 55% for gradient, PNG and self-similarity measures, respectively) and the lowest

in PPA (11% for the gradient measure), LOC (12% for the PNG measure) or FFA (7% for the self-similarity measure). Furthermore, for all of the investigated visual complexity measures, the encoding performance was higher in the ROIs in the lower visual cortex and relatively low in the ROIs in the higher visual cortex and V4. Compared to the gradient measure, the stimulus-evoked voxel responses in the PPA region could be better predicted from the PNG and self-similarity measures.

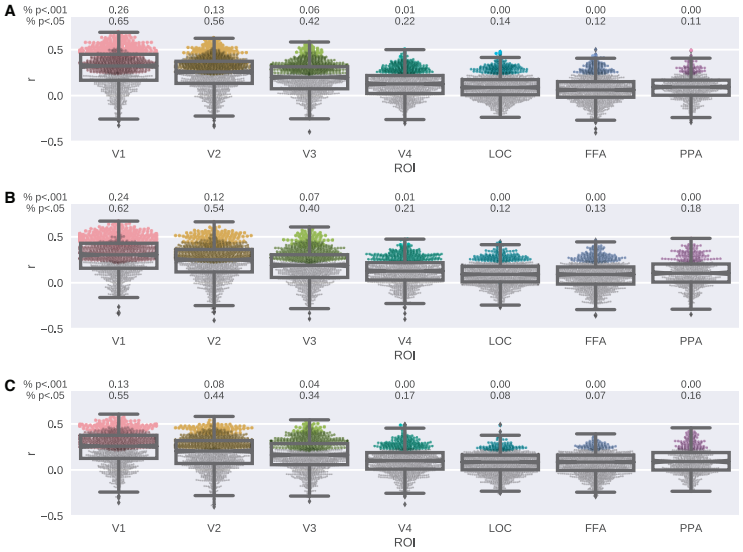

Figure 4.2: Experiment 1 - Distributions of encoding performance over individual voxels in visual ROIs. ***A***: Gradient. ***B***: PNG. ***C***: Self-similarity. Boxes show interquartile range. Notches show second quartile. Whiskers show ± 1.5 interquartile range. Points show encoding performance of individual voxels. Colors show p-values (black: outlier; gray: $p \geq 0.05$; dark: $p < 0.05$; light: $p < 0.001$).

Next, we investigated the overlap between different visual complexity measures in terms of the number of voxels whose responses were significantly predicted in the lower and higher visual cortices (Figure 4.3). In the lower visual cortex, the amount of overlap between the significantly predicted voxel responses of all three visual complexity measures was relatively high with 46% of all

significant voxels. This overlap reduced to 20% in ROIs in the higher visual cortex and V4. Furthermore, the amount of overlap between the gradient and PNG measures was very high (73%) in the lower visual cortex. This overlap reduced to 44% in the ROIs in the higher visual cortex. Out of all the significantly predicted voxels in the higher visual cortex, 17% were only sensitive to the gradient measure, whereas this number was 13% for the PNG measure and 17% for the self-similarity measure.

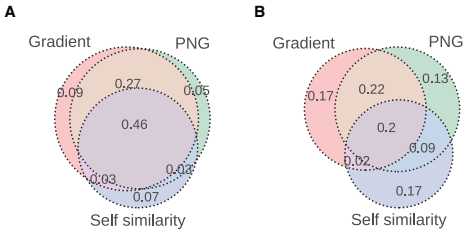

Figure 4.3: Experiment 1 - Overlap between different visual complexity measures in the visual cortex. ***A***: Ratio of overlapping voxels whose responses were significantly predicted in the lower visual cortex. ***B***: Ratio of overlapping voxels whose responses were significantly predicted in the higher visual cortex and V4.

Finally, we investigated the sensitivity of the voxels whose responses were significantly predicted ($p < 0.05$) in each ROI to the measures of visual complexity by calculating the mean absolute beta coefficient (i.e. slope) of the regression models for each ROI. The beta coefficients show the extent to which the response of a voxel is modulated per one standard deviation change in the complexity value of the stimulus. Figure 4.4 shows the results of these analyses. We observed that while in V1 the mean slope was the highest, it decreased almost gradually as we moved to higher visual areas, suggesting that the representations of complexity became coarser along the visual hierarchy. Since this pattern was very similar to the encoding performance, we further calculated the slopes controlled for the encoding performance by dividing the beta coefficients by the corresponding correlation coefficients (Panel B in Figure 4.4). This did not change the observed pattern,

confirming that the slopes of betas were indeed indicative of the sensitivity of complexity representations of voxels rather than just reflecting the encoding performance. Another interesting result that we found was that large portions of the voxels in all ROIs of the visual cortex had negative slopes, such that their response increased as the complexity of the image decreased. While a majority of the voxels in the lower visual areas and V4 had positive slopes, in the higher visual areas voxels selective to simplicity rather than complexity were more common.

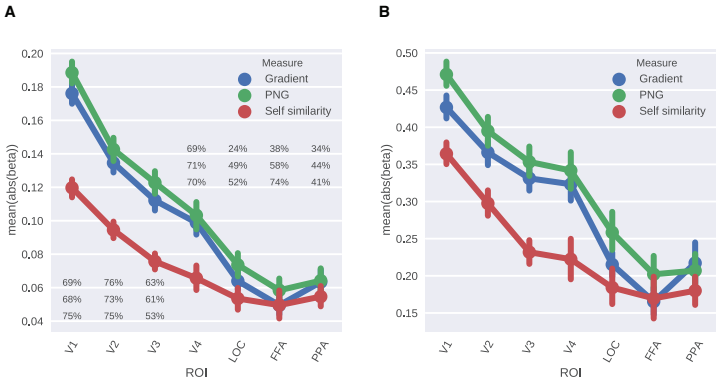

Figure 4.4: Experiment 1 - ***A***: Mean absolute beta coefficients over the significant voxels in the visual ROIs. Percentages show the percentage of positive beta coefficients corresponding to each complexity measure: Gradient, PNG and Self-similarity, from top to bottom, respectively. ***B***: Mean absolute normalized beta (beta / r) over the significant voxels in the visual ROIs. Error bars show ± 1 SEM.

Experiment 2

First, to understand the relationship between the auditory complexity measures, we calculated bivariate Pearson correlation coefficients between each measure (Table 4.6). All three measures were significantly correlated with each other ($p < 0.05$). Highest correlation was between the two compression measures FLAC and Ogg ($r = 0.66$), whereas the lowest correlation was observed between event density and Ogg ($r = 0.23$).

Table 4.6: Experiment 2 - Bivariate correlations (r) between auditory complexity measures.

	Event density	FLAC	Ogg
Event density	1.00	0.45	0.23
FLAC	0.45	1.00	0.66
Ogg	0.23	0.66	1.00

Next, we performed the decoding analysis on the hyperaligned data of the eight subjects. That is, for each one of the auditory complexity measures, we predicted the value of the complexity measure from the stimulus-evoked responses of voxels in the ROIs from the early auditory and auditory association cortices (Table 4.2) using ridge regression. The results of these analyses are presented in Figure 4.5. The most striking (but not unexpected) result was that for the two file compression measures FLAC and Ogg, the general pattern of the decoding performances in ROIs were very similar with each other, but not very similar to the event density measure. For the event density measure, decoding performance was worse than those of the compression measures in all ROIs except for the STSda region. Overall, the highest decoding performance was obtained for the FLAC measure with the highest correlation results in the PBelt region ($r = 0.74$, $p \ll 0.001$) followed by MBelt ($r = 0.70$, $p \ll 0.001$) and A1 ($r = 0.70$, $p \ll 0.001$) regions. For the two compression measures, all ROIs in the early auditory cortices were decoded significantly above chance, with all r values above 0.57. In the early auditory cortex, largest differences between the decoding performance were observed in the RI region between the event density and both of the compression measures, where for event density the predictions were not better than chance and for FLAC and Ogg, correlations were rather high ($r = 0.69$, $p \ll 0.001$ and $r = 0.59$, $p \ll 0.001$ for FLAC and Ogg, respectively). Conversely, in STSda, neither FLAC nor Ogg measure could be predicted significantly above the chance level, whereas event density was

decoded significantly ($r = 0.19$, $p < 0.05$). Among the ROIs in the auditory association cortex, the compression measures could be best predicted from the voxels in A4 ($r = 0.62$, $p \ll 0.001$ and $r = 0.49$, $p \ll 0.001$ for FLAC and Ogg, respectively) and TA2 ($r = 0.53$, $p \ll 0.001$ and $r = 0.44$, $p \ll 0.001$ for FLAC and Ogg, respectively) regions. The only ROI that none of the auditory complexity measures could be predicted from was STGa.

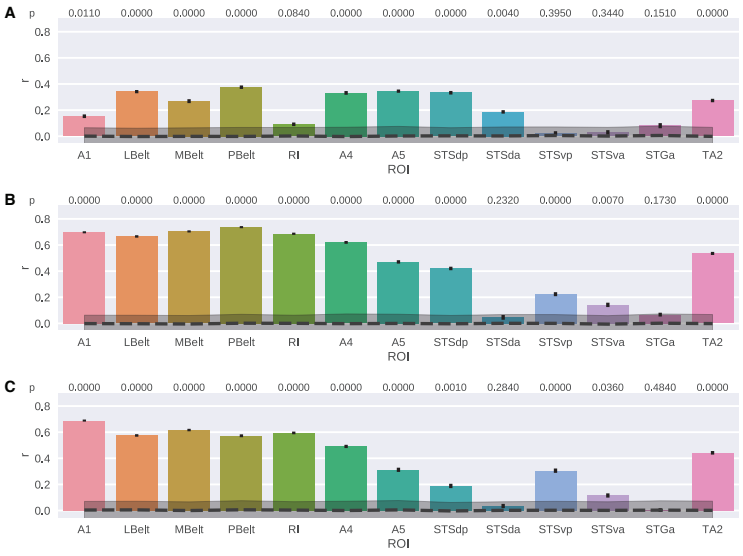

Figure 4.5: Experiment 2 - Results of decoding auditory complexity measures from the ROIs in the auditory cortex. ***A***: Event density. ***B***: FLAC. ***C***: Ogg. Bars and error bars show decoding performance and ± 1 SEM, respectively. Dashed line and shaded region show chance level and ± 1 SEM around $r = 0$, respectively. SEM is computed with bootstrapping (1000 iterations).

Following the decoding analyses, we performed encoding analysis, again on the hyperaligned data of the eight subjects, to evaluate how well the stimulus-evoked responses in ROIs could be predicted from the auditory complexity measures using linear regression. The results of the encoding analyses are presented in Figure 4.6. Similar to the decoding results, the general pattern of the encoding performances of the two compression measures were very high and resembled each other. A large majority of the

voxel responses in the early auditory cortices could be predicted significantly above chance level ($p < 0.05$) using FLAC and Ogg measures. The percentage of the significantly predicted voxel responses ranged from 72% to 91% for the FLAC measure and from 71% to 87% for the Ogg measure. Among the ROIs in the auditory associative cortex best encoding performance was in A5 for FLAC and A4 for Ogg, whereas the worst performance was in STSva for both compression measures. The encoding performance of the regression models using the event density measure was relatively low in all ROIs with the highest percentage of significantly above chance level ($p < 0.05$) predictions in PBelt (43%), A4 (34%) and TA2 (41%) regions.

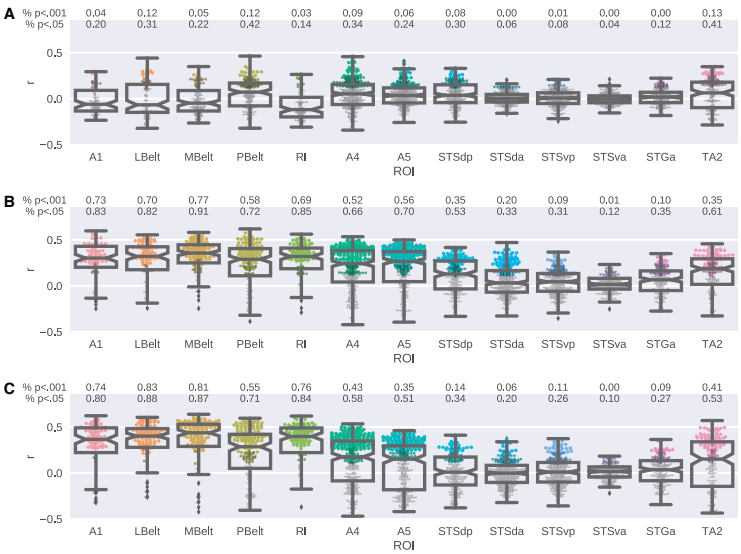

Figure 4.6: Experiment 2 - Distributions of encoding performance over individual voxels in auditory ROIs. ***A***: Event density. ***B***: FLAC. ***C***: Ogg. Boxes show interquartile range. Notches show second quartile. Whiskers show ± 1.5 interquartile range. Points show encoding performance of individual voxels. Colors show p-values (black: outlier; gray: $p \geq 0.05$; dark: $p < 0.05$; light: $p < 0.001$).

Next, we investigated the overlap between different auditory complexity measures in terms of the number of voxels whose

responses were significantly predicted in the early auditory and auditory association cortices (Figure 4.7). The amount of overlap between the significantly predicted voxel responses of FLAC and Ogg measures was very high (77%) in the early auditory cortex. This overlap reduced to 47% in the ROIs in the auditory associative cortex. Out of all the significantly predicted voxels in the auditory associative cortex, 29% were only sensitive to the FLAC measure, whereas this number was 10% for the Ogg measure and 12% for the event density. The voxels that were significantly predicted by all three measures made up 15% of all significant voxels in the early auditory cortex and 11% in the auditory associative cortex.

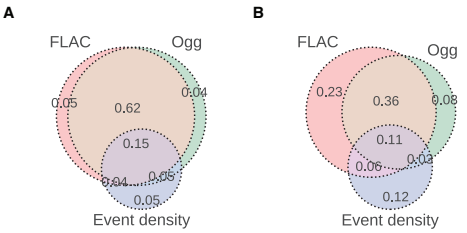

Figure 4.7: Experiment 2 - Overlap between different auditory complexity measures in the auditory cortex. ***A***: Ratio of overlapping voxels whose responses were significantly predicted in the early auditory cortex. ***B***: Ratio of overlapping voxels whose responses were significantly predicted in the auditory association cortex.

Finally, we looked at how the mean beta coefficients varied between different ROIs (Figure 4.8). We observed that there was a large difference between the early auditory cortex and the auditory association cortex for the FLAC and Ogg measures, such that the early regions had a finer sensitivity to complexity and the associative regions had a coarser sensitivity to changes in complexity levels. For the event density measure, the mean beta levels were low in all ROIs. When controlled for the encoding performance (Panel B in Figure 4.8) the differences between the early and associative auditory cortices remained similar except for in the STSva region for the Ogg measure which had a large vari-

ability. This analysis also revealed that the STSda region showed high sensitivity to the event density measure. Overall, the mean betas of the event density measure showed relatively larger variability both between and within ROIs. Regarding the direction of the relationship between complexity and voxel responses, for the compression complexity measures FLAC and Ogg (unlike in the visual cortex) the majority of voxels in all auditory ROIs had positive beta coefficients, i.e they responded more as the stimulus complexity increased. However for the event density measure, most voxels had negative slopes.

Figure 4.8: Experiment 2 - ***A***: Mean absolute beta coefficients over the significant voxels in the auditory ROIs. Percentages show the percentage of positive beta coefficients corresponding to each complexity measure: Event density, FLAC and Ogg, from top to bottom, respectively. ***B***: Mean absolute normalized beta (beta / r) over the significant voxels in the auditory ROIs. Error bars show ± 1 SEM.

4.4 Discussion

In this study, we investigated the neural representations of image and music complexity in the human visual and auditory cortices. To this end, we performed univariate encoding and multivariate decoding analyses of fMRI data from two different experiments measuring the stimulus-evoked BOLD responses to large collec-

tions of photograph and music stimuli. In Experiment 1, we found that visual complexity was represented throughout the visual cortex with decreasing sensitivity from lower to higher visual areas. While all regions in the lower visual cortex were highly responsive to stimulus complexity, in the higher visual cortex, PPA was the most responsive region to visual complexity of photographs. Voxels representing complexity in the lower visual cortex mostly showed increased activity with increased complexity, whereas approximately half of those in the higher visual areas showed decreased activity as image complexity increased. In the case of representations of auditory complexity, in Experiment 2, we found that encoding and decoding performances and sensitivity of voxels were high for the early auditory cortex and on average lower for the regions in the auditory association cortex. Among the ROIs in the auditory association cortex, A4 and TA2 were the most responsive regions to stimulus complexity. Furthermore, we determined that representations of event density in music were less pronounced compared to those of Kolmogorov complexity measures throughout the auditory cortex.

The differences between the neural representations of different complexity measures, which are also suggested by varying degrees of correlations between these measures, became more pronounced after a fine-grained analysis of the results. In both visual and auditory cortices, we showed that in the earlier sensory areas, the overlap between voxels representing different complexity measures were higher than the overlap in the areas higher in the sensory hierarchy. This result implies that voxels coding for different complexity dimensions become more specialized along the visual and auditory hierarchy.

Furthermore, both in the auditory and the visual cortices, the sensitivity of voxels to changes in complexity – measured by the magnitudes of the regression slopes of the voxel encoding models – decreased along the sensory hierarchy. While this decrease was gradual in the visual cortex, it was more abrupt starting

at the A4 region in the auditory association cortex, and in the auditory cortex was only observed for the Kolmogorov complexity measures. These results are reminiscent of the representational gradients of other sensory stimulus features such as semantic features of movies(Huth, Nishimoto, Vu & Gallant, 2012) and speech(Huth, de Heer, Griffiths, Theunissen & Gallant, 2016), and task-optimized features of images(Güçlü & van Gerven, 2015), movies(**Guclu2017**) or music(Güçlü et al., 2016), which further suggest a functional organization in terms of gradients rather than patches.

Another difference between the complexity representations in the two sensory cortices was observed between the number of voxels showing increased activity with increased complexity. In the auditory cortex, the responses of voxels were mostly positively correlated with complexity, whereas in the higher visual cortex, around half of the voxels had negative beta coefficients indicating increased activity in response to increased stimulus simplicity rather than complexity.

We found that from PPA, all of the tested complexity measures could be decoded with a very high accuracy but this was not the case for the encoding analysis. We believe that this suggests the presence of distributed representations of complexity in PPA (cf. Park et al. (Park, Brady, Greene & Oliva, 2011)) rather than single voxels encoding for information regarding scene complexity. Possibly, our multivariate decoding approach allowed us to make accurate predictions about the complexity of the stimulus, whereas our univariate encoding analysis did not allow to capture the distributed representations of complexity.

The decoding accuracy in PPA was on par with the regions in the early visual cortex and was much better than other higher level regions such as FFA and LOC. This result might seem surprising at first glance, however it is actually not unexpected given that PPA is primarily responsible for representing scenes (Epstein &

Kanwisher, 1998; Epstein, 2005), for which complexity is one of the defining properties that allows identifying different scene categories (Oliva & Torralba, 2001). On the other hand, such a relationship has not been established to the same extent for objects and faces. Our results are in line with the accepted functional role of the PPA in scene processing and support previous behavioural results showing that scene identification utilizes global image properties (Renninger & Malik, 2004; Greene & Oliva, 2010).

The auditory association regions that we identified to represent music complexity largely overlapped with the regions that have been shown to activate during story listening and auditory math tasks, whereas we observed no strong correspondences with the story - math contrast (Glasser et al., 2016). As seen in Figure **??** decoding of complexity levels from anterior STS and STG regions were rather unsuccessful, i.e. these regions performed either not or merely significantly different from chance. For example, in STSda, only event density could be significantly decoded and this was with a low amount of correlation. Similarly in STSva only FLAC and Ogg could be decoded with a performance that was merely significantly above chance and no measure could be predicted from STGa. Moving from anterior to posterior regions, the decoding performance of the ROIs increased, such that performance in both STSvp and STSdp was significant, and in A4 and A5 regions, which lie on posterior STG, it was the highest among the areas in the auditory association cortex. (Left) posterior STG and STS regions have been shown to process syntactic complexity of language in several previous studies (for a review, see(Friederici, 2011)). In most of these studies syntactic complexity is investigated by comparing list of words to sentences, sentences with simple syntactic structures to those with complex ones(Friederici, 2011). Our results show that posterior STG and STS regions not only process syntactic complexity, but also they represent music complexity.

In the auditory association cortex, besides the ROIs on posterior STG and STS, we identified TA2 (which lies on planum polare) as a region that represents complexity well. In terms of speech and other complex sounds, the function of this region has not been well established, however it is known to show greater activity in response to music, compared to vocal and speech sounds(**Samson2011**). Our results demonstrate that TA2 is a region which also represents complexity of music.

Event density has been suggested as a good measure of the complexity of songs (Marin & Leder, 2013; Bader, 2013). However, based on our results, event density showed different neural representations than the Kolmogorov complexity measures in the auditory cortex. Furthermore, event density was less well encoded and decoded by our approaches compared to the FLAC and Ogg measures.

We have used three different complexity measures per sensory modality so as to not overlook the multifaceted nature of complexity as discussed in the Introduction section. Both in the visual and auditory domains, this multifaceted nature of complexity makes it difficult to perfectly define complexity and select a "best" measure, especially when estimating the complexity levels of naturalistic stimuli. By reporting the results of our analysis for a selection of computational complexity measures, we aimed to provide a greater insight into how different measures of complexity are processed in the human brain. We hope that the differences among the different measures that we report here will be a useful resource for future studies.

In this study, our goal was to establish a direct, predictive relationship between objective stimulus complexity and stimulus-evoked brain activations at single subject level rather than making inferences at population level in accordance with previous neural encoding and decoding studies in the literature (Naselaris et al., 2011; Kay, Naselaris, Prenger & Gallant, 2008; Mitchell et al.,

2008; Güçlü & van Gerven, 2015). As such, we used data sets with a very large number of data points per subject (approximately 3 h and 9 h per subject) but a small number of subjects in total (eight and five subjects in total) for training and testing predictive models on separate datasets. Therefore, we performed statistical analyses and tests at single subject level. As a result, we were able to establish such a direct, predictive relationship. However, it should be noted that making inferences at population level would require a larger group study.

Literature in art perception suggests a strong link between complexity and aesthetic responses. A previous study investigating aesthetic responses to mathematical formulae indirectly provides some insight into how elegance (or simplicity) of highly intellectual and abstract concepts are processed in the brain(Zeki, Romaya, Benincasa & Atiyah, 2014). The authors of the study found that the beautiful mathematical formulae (which in many case were the formulae which were simple yet meaningful) activated the A1 region of the medial orbito-frontal cortex, which is known also to activate in response to beauty of art. However, this study investigated conceptual or mathematical elegance rather than perceptual simplicity. Therefore, it would still be interesting to investigate how stimulus complexity manifests itself throughout the brain (besides the currently investigated sensory cortices) in response to artworks, images and music of varying complexity levels. Moreover, an interesting next step would be to investigate the effects of additional factors such as scene preferences(Martínez-Soto, Gonzales-Santos, Pasaye & Barrios, 2013) on the neural representations of stimulus complexity. Finally, while all of the objective computational complexity measures such as Kolmogorov complexity (PNG, FLAC and Ogg), gradient, self-similarity and event density that were employed in this study are established estimates of subjective complexity of auditory and visual stimuli, subjectively rated stimulus complexity can be investigated in future studies.

Supplementary Materials

Image Categories

Category 1: French horn, horn. Category 2: Frisbee. Category 3: Kalashnikov. Category 4: Segway, Segway Human Transporter, Segway HT. Category 5: airliner. Category 6: airship, dirigible. Category 7: baby buggy, baby carriage, carriage, perambulator, pram, stroller, go-cart, pushchair, pusher. Category 8: backpack, back pack, knapsack, packsack, rucksack, haversack. Category 9: barrow, garden cart, lawn cart, wheelbarrow. Category 10: baseball bat, lumber. Category 11: baseball glove, glove, baseball mitt, mitt. Category 12: basket, basketball hoop, hoop. Category 13: bat, chiropteran. Category 14: bathtub, bathing tub, bath, tub. Category 15: beacon, lighthouse, beacon light, pharos. Category 16: bear. Category 17: beer mug, stein. Category 18: binoculars, field glasses, opera glasses. Category 19: birdbath. Category 20: bonsai. Category 21: bowling ball, bowl. Category 22: bowling pin, pin. Category 23: boxing glove, glove. Category 24: bulldozer, dozer. Category 25: butterfly. Category 26: camel. Category 27: camera tripod. Category 28: cannon. Category 29: canoe. Category 30: car wheel. Category 31: centipede. Category 32: cereal box. Category 33: chimpanzee, chimp, Pan troglodytes. Category 34: clasp knife, jackknife. Category 35: cockroach, roach. Category 36: coffee mug. Category 37: coffin, casket. Category 38: common iguana, iguana, Iguana iguana. Category 39: common raccoon, common racoon, coon, ringtail, Procyon lotor. Category 40: compact disk, compact disc, CD. Category 41: computer keyboard, keypad. Category 42: computer monitor. Category 43: conch. Category 44: cormorant, Phalacrocorax carbo. Category 45: covered wagon, Conestoga wagon, Conestoga, prairie wagon, prairie schooner. Category 46: cowboy hat, ten-gallon hat. Category 47: crab. Category 48: dial telephone, dial phone. Category 49: diskette, floppy, floppy disk.

Category 50: dog, domestic dog, Canis familiaris. Category 51: dolphin. Category 52: domestic llama, Lama peruana. Category 53: dress hat, high hat, opera hat, silk hat, stovepipe, top hat, topper, beaver. Category 54: duck. Category 55: dumbbell. Category 56: earphone, earpiece, headphone, phone. Category 57: electric guitar. Category 58: elephant. Category 59: elk, European elk, moose, Alces alces. Category 60: fighter, fighter aircraft, attack aircraft. Category 61: fire engine, fire truck. Category 62: fire extinguisher, extinguisher, asphyxiator. Category 63: flashlight, torch. Category 64: football helmet. Category 65: frog, toad, toad frog, anuran, batrachian, salientian. Category 66: frying pan, frypan, skillet. Category 67: gas pump, gasoline pump, petrol pump, island dispenser. Category 68: ghetto blaster, boom box. Category 69: giraffe, camelopard, Giraffa camelopardalis. Category 70: goat, caprine animal. Category 71: goldfish, Carassius auratus. Category 72: golf ball. Category 73: goose. Category 74: gorilla, Gorilla gorilla. Category 75: grand piano, grand. Category 76: grape. Category 77: grasshopper, hopper. Category 78: gravestone, headstone, tombstone. Category 79: greyhound. Category 80: guitar pick. Category 81: gym shoe, sneaker, tennis shoe. Category 82: hammock, sack. Category 83: hand calculator, pocket calculator. Category 84: harmonica, mouth organ, harp, mouth harp. Category 85: harp. Category 86: harpsichord, cembalo. Category 87: hawksbill turtle, hawksbill, hawkbill, tortoiseshell turtle, Eretmochelys imbricata. Category 88: helicopter, chopper, whirlybird, eggbeater. Category 89: homo, man, human being, human. Category 90: horse, Equus caballus. Category 91: horseshoe crab, king crab, Limulus polyphemus, Xiphosurus polyphemus. Category 92: hot tub. Category 93: hot-air balloon. Category 94: hourglass. Category 95: housefly, house fly, Musca domestica. Category 96: hummingbird. Category 97: iPod. Category 98: ibis. Category 99: joystick. Category 100: kangaroo. Category 101: kayak. Category 102: ketch. Category 103: killer whale, killer, orca, grampus, sea wolf, Orcinus orca. Category 104: knife. Category 105: knob, boss. Category 106: laptop, laptop computer. Category 107: lathe. Category 108: lawn mower, mower. Cat-

egory 109: leopard, Panthera pardus. Category 110: light bulb, lightbulb, bulb, incandescent lamp, electric light, electric-light bulb. Category 111: mailbox, letter box. Category 112: mandolin. Category 113: marimba, xylophone. Category 114: megaphone. Category 115: menorah. Category 116: microscope. Category 117: microwave, microwave oven. Category 118: minaret. Category 119: miniature fan palm, bamboo palm, fern rhapis, Rhapis excelsa. Category 120: motorcycle, bike. Category 121: mountain bike, all-terrain bike, off-roader. Category 122: mouse, computer mouse. Category 123: mushroom. Category 124: mussel. Category 125: necktie, tie. Category 126: obelisk. Category 127: octopus, devilfish. Category 128: ostrich, Struthio camelus. Category 129: owl, bird of Minerva, bird of night, hooter. Category 130: palm, palm tree. Category 131: paper clip, paperclip, gem clip. Category 132: penguin. Category 133: personal digital assistant, PDA, personal organizer, personal organiser, organizer, organiser. Category 134: photocopier. Category 135: pincer, pair of pincers, tweezer, pair of tweezers. Category 136: pitcher, ewer. Category 137: planchet, coin blank. Category 138: pool table, billiard table, snooker table. Category 139: porcupine, hedgehog. Category 140: praying mantis, praying mantid, Mantis religioso. Category 141: projector. Category 142: radio telescope, radio reflector. Category 143: refrigerator, icebox. Category 144: revolver, six-gun, six-shooter. Category 145: rifle. Category 146: roulette wheel, wheel. Category 147: saddle. Category 148: school bus. Category 149: scorpion. Category 150: screwdriver. Category 151: sextant. Category 152: shirt. Category 153: shredder. Category 154: skateboard. Category 155: skunk, polecat, wood pussy. Category 156: skyscraper. Category 157: snail. Category 158: snake, serpent, ophidian. Category 159: snowmobile. Category 160: soccer ball. Category 161: sock. Category 162: soda can. Category 163: spectacles, specs, eyeglasses, glasses. Category 164: speedboat. Category 165: spider. Category 166: spoon. Category 167: stained-glass window. Category 168: starfish, sea star. Category 169: steering wheel, wheel. Category 170: stirrup, stirrup iron. Category 171: sunflower, helianthus. Category 172: swan.

Category 173: sword, blade, brand, steel. Category 174: syringe. Category 175: tambourine. Category 176: teapot. Category 177: telephone booth, phone booth, call box, telephone box, telephone kiosk. Category 178: tennis ball. Category 179: tennis racket, tennis racquet. Category 180: tepee, tipi, teepee. Category 181: theodolite, transit. Category 182: toaster. Category 183: toaster oven. Category 184: tomato. Category 185: treadmill. Category 186: triceratops. Category 187: tricycle, trike, velocipede. Category 188: trilobite. Category 189: true toad. Category 190: tuning fork. Category 191: umbrella. Category 192: videocassette recorder, VCR. Category 193: washer, automatic washer, washing machine. Category 194: watch, ticker. Category 195: watermelon. Category 196: welder's mask. Category 197: windmill. Category 198: wine bottle. Category 199: yarmulke, yarmulka, yarmelke. Category 200: zebra.

Music Tags

Tag 1: airy. Tag 2: ambient. Tag 3: arabic. Tag 4: beats. Tag 5: bells. Tag 6: blues. Tag 7: cello. Tag 8: classical. Tag 9: country. Tag 10: dance. Tag 11: dark. Tag 12: drums. Tag 13: eastern. Tag 14: eerie. Tag 15: electric. Tag 16: electric guitar. Tag 17: electro. Tag 18: fast. Tag 19: female singing. Tag 20: flutes. Tag 21: guitars. Tag 22: hard rock. Tag 23: harp. Tag 24: harpsichord. Tag 25: heavy metal. Tag 26: horns. Tag 27: indian. Tag 28: instrumental. Tag 29: jazz. Tag 30: jazzy. Tag 31: loud. Tag 32: man singing. Tag 33: metal. Tag 34: middle eastern. Tag 35: new age. Tag 36: no guitars. Tag 37: no singing. Tag 38: noise. Tag 39: opera. Tag 40: oriental. Tag 41: piano. Tag 42: pop. Tag 43: quiet. Tag 44: rock. Tag 45: sax. Tag 46: singing. Tag 47: sitar. Tag 48: slow. Tag 49: soft. Tag 50: soft rock. Tag 51: strange. Tag 52: strings. Tag 53: synth. Tag 54: techno. Tag 55: trumpet. Tag 56: violins. Tag 57: weird. Tag 58: wind. Tag 59: world.

5 Convolutional Sketch Inversion

This chapter is based on Güçlütürk, Y., Güçlü, U., van Lier, R., and van Gerven, M. (2016). Convolutional sketch inversion. In Computer Vision for Art Analysis - European Conference on Computer Vision. https://doi.org/10.1007/978-3-319-46604-0$_56$

5.1 Introduction

Portrait and self-portrait sketches have an important role in art. From an art historical perspective, self-portraits serve as historical records of what the artists looked like. From the perspective of an artist, self-portraits can be seen as a way to practice and improve one's skills without the need for a model to pose. Portraits of others further serve as memorabilia and a record of the person in the portrait. Artists most often are able to easily capture recognizable features of a person in their sketches. Therefore, hand-drawn sketches of people have further applications in law enforcement. Sketches of suspects drawn based on eye-witness accounts are used to identify suspects, either in person or from catalogues of mugshots.

Prior work related to face sketches in computer vision has been mostly limited to synthesis of highly controlled (i.e. having neutral expression, frontal pose, with normal lighting and without any occlusions) sketches from photographs, i.e. sketch synthesis (Tang & Wang, 2003; Liu, Tang, Jin, Lu & Ma, 2005; Gao, Wang, Tao & Li, 2012; Wang, Tao, Gao, Li & Li, 2013b; Zhang, Lin, Wu, Ding & Zhang, 2015) and photographs from sketches, i.e. sketch inversion (Liu, Tang & Liu, 2007; Xiao, Gao, Tao & Li, 2009; Wang & Tang, 2009; Gao et al., 2012; Wang et al., 2013b). Sketch inversion studies with controlled inputs utilized patch-based approaches and used Bayesian tensor inference (Liu et al., 2007), an embedded hidden Markov model (Xiao et al., 2009), a multiscale Markov random field model (Wang & Tang, 2009), sparse representations (Gao et al., 2012) and transductive learning with a probabilistic graph model (Wang et al., 2013b).

Few studies developed methods of sketch synthesis to handle more variation in one or more variables at a time, such as lighting (Li, Savvides & Bhagavatula, 2006), and lighting and pose (Zhang, Wang & Tang, 2010). In a recent study, Zhang, Gao, Wang

and Li (2016) showed that sketch synthesis by transferring the style of a single sketch could be used also in uncontrolled conditions. In (Zhang et al., 2016), first an initial sketch by a sparse representation-based greedy search strategy was estimated, then candidate patches were selected from a template style sketch and the estimated initial sketch. Finally, the candidate patches were refined by a multi-feature-based optimization model and the patches were assembled to produce the final synthesized sketch.

Recently, the use of deep convolutional neural networks (DNNs) in image transformation tasks, in which one type of image is transformed into another, has gained tremendous traction. In the context of sketch analysis, DNNs were used to tackle the problems of sketch synthesis and sketch simplification. For example, Zhang et al. (2015) has used a DNN to convert photographs to sketches. They developed a DNN with six convolutional layers and a discriminative regularization term for enhancing the discriminability of the generated sketch against other sketches. Furthermore, Simo-Serra, Iizuka, Sasaki and Ishikawa (2016) has used a DNN to simplify rough sketches. They have shown that users prefer sketches simplified by the DNN more than they do those by other applications 97% of the time.

Some other notable image transformation problems include colorization, style transfer and super-resolution. In colorization, the task is to transform a grayscale image to a color image that accurately captures the color information. In style transfer, the task is to transform one image to another image that captures the style of a third image. In super-resolution, the task is to transform a low-resolution image to a high-resolution image with maximum quality. DNNs have been used to tackle all of these problems with state-of-the art results (Cheng, Yang & Sheng, 2015; Iizuka, Simo-Serra & Ishikawa, 2016; Gatys, Ecker & Bethge, 2015; Dong, Loy, He & Tang, 2014, 2016; Johnson, Alahi & Li, 2016).

However, a challenging task that remains is photorealistic face image synthesis from face sketches in uncontrolled conditions. That is, at present, there exist no sketch inversion models that are able to perform in realistic conditions. These conditions are characterized by changes in expression, pose, lighting condition and image quality, as well as the presence of varying amounts of background clutter and occlusions.

Here, we use DNNs to tackle the problem of inverting face sketches to synthesize photorealistic face images from different sketch styles in uncontrolled conditions. We developed three different models to handle three different types of sketch styles by training DNNs on datasets that we constructed by extending a well-known large-scale face dataset, obtained in uncontrolled conditions (Liu, Luo, Wang & Tang, 2015). We test the models on another similar large-scale dataset (Learned-Miller, Huang, RoyChowdhury, Li & Hua, 2016), a hand-drawn sketch database (Wang & Tang, 2009) as well as on self-portrait sketches of famous Dutch artists. We show that our approach, which we refer to as *Convolutional Sketch Inversion* (CSI) can be used to achieve state-of-the-art results and discuss possible applications in fine arts, art history and forensics.

5.2 Materials and Methods

Semi-Simulated Datasets

For training and testing our CSI model, we made use of the following datasets:

- *Large-scale CelebFaces Attributes (CelebA) dataset* (Liu et al., 2015). The CelebA dataset contains 202,599 celebrity face images and 10,177 identities. The images were obtained from the internet and vary extensively in terms of pose,

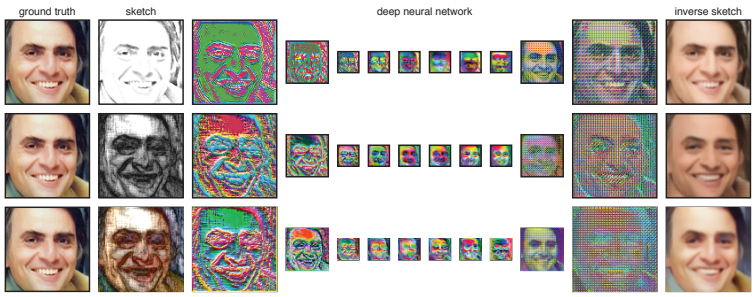

Figure 5.1: Example results of our convolutional sketch inversion models. Our models invert face sketches to synthesize photorealistic face images. Each row shows the sketch inversion / photo synthesis pipeline that transforms a different sketch of the same face to a different image of the same face via a different deep neural network. Each deep neural network layer is represented by the top three principal components of its feature maps.

expression, lighting, image quality, background clutter and occlusion. Each image in the dataset has five landmark positions and 40 attributes. These images were used for training the networks.

- *Labeled Faces in the Wild (LFW) dataset* (Learned-Miller et al., 2016). The LFW dataset contains 13,233 face images and 5749 identities. Similar to the CelebA dataset, images were obtained from the internet and vary extensively in terms of pose, expression, lighting, image quality, background clutter and occlusion. A subset of these images (11,990) were used for testing the networks.

- *CUHK Face Sketch (CUFS) database* (Wang & Tang, 2009). The CUFS database contains photographs and their corresponding hand-drawn sketches of 606 individuals. The dataset was formed by combining face photographs from three other databases and producing hand-drawn sketches of these photographs. Concretely, it consists of 188 face photographs from the Chinese University of Hong Kong

(CUHK) student database (Wang & Tang, 2009) and their corresponding sketches, 123 face photographs from the AR Face Database (Martinez & Benavente, 1998) and their corresponding sketches, and 295 face photographs from the XM2VTS database (Messer, anad J. Kittler, Luettin & Maître, 1999) and their corresponding sketches. Only 18 of the sketches (six from each sub-database) were used in the current study. These images were used for testing the networks.

- *Sketches of famous Dutch artists*. We also used the following sketches: i) Self-Portrait with Beret, Wide-Eyed by Rembrandt, 1630, etching, ii) Two Self-portraits and Several Details by Vincent van Gogh, 1886, pencil on paper and iii) Self-Portrait by M.C. Escher, 1929, lithograph on gray paper. These images were used for testing the networks.

Preprocessing

Similar to Cowen et al. (2014), each image was cropped and resized to 96 pixels $\times$ 96 pixels such that:

- The distance between the top of the image and the vertical center of the eyes was 38 pixels.
- The distance between the vertical center of the eyes and the vertical center of the mouth was 32 pixels.
- The distance between the vertical center of the mouth and the bottom of the image was 26 pixels.
- The horizontal center of the eyes and the mouth was at the horizontal center of the image.

Sketching

Each image in the CelebA and LFW datasets was automatically transformed to a line sketch, a grayscale sketch and a color sketch. Sketches in the CUFS database and those by the famous Dutch artists were further transformed to line sketches by using the same procedure.

Color and grayscale sketch types are produced by the same stylization algorithm (Gastal & Oliveira, 2011). To obtain the sketch images, the input image is first filtered by an edge-aware filter. This filtered image is then blended with the magnitude of the gradient of the filtered image. Then, each pixel is scaled by a normalization factor resulting in the final sketch-like image.

Line sketches which resemble pencil sketches were generated based on Beyeler (2015). Line sketch conversion works by first converting the color image to grayscale. This is followed by inverting the grayscale image to obtain a negative image. Next, a Gaussian blur is applied. Finally, using color dodge, the resulting image is blended with the grayscale version of the original image.

It should be noted that synthesizing face images from color or grayscale sketches is a more difficult problem than doing so from line sketches since many details of the faces are preserved by line sketches while they are lost for other sketch types.

Models

We developed one DNN for each of the three sketch styles based on the style transfer architecture in (Johnson et al., 2016). Each of the three DNNs was based on the same architecture except for the first layer where the number of input channels were either one or three depending on the number of color channels of the sketches.

The architecture comprised four convolutional layers, five residual blocks (He, Zhang, Ren & Sun, 2015), two deconvolutional layers and another convolutional layer. Each of the five residual blocks comprised two convolutional layers. All of the layers except for the last layer were followed by batch normalization (Ioffe & Szegedy, 2015) and rectified linear units. The last layer was followed by batch normalization and hyperbolic tangent units. All models were implemented in the Chainer framework (Tokui, Oono, Hido & Clayton, 2015). Table 5.1 shows the details of the architecture.

Table 5.1: Deep neural network architectures. BN; batch normalization with decay = 0.9, $\epsilon = 1e-5$, ReLU; rectified linear unit, con.; convolution, dec.; deconvolution, res.; residual block, tanh; hyperbolic tangent unit. Outputs of the hyperbolic tangent units are scaled to [0, 255]. x/y indicates the parameters of the first and second layers of a residual block. +x indicates that the input and output of a block are summed and no activation function is used.

Layer	Type	in_channels	out_channels	ksize	stride	pad	normalization	activation
1	con.	1 or 3	32	9	1	4	BN	ReLU
2	con.	32	64	3	2	1	BN	ReLU
3	con.	64	128	3	2	1	BN	ReLU
4	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
5	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
6	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
7	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
8	res.	128/128	128/128	3/3	1/1	1/1	BN/BN	ReLU/+x
9	dec.	128	64	3	2	1	BN	ReLU
10	dec.	64	32	3	2	1	BN	ReLU
11	con.	32	3	9	1	4	BN	tanh

Estimation

For model optimization we used Adam (Kingma & Ba, 2014) with parameters $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_=0.999$, $\epsilon = 10^{-8}$ and mini-batch size = 4. We trained the models by iteratively minimizing the loss function for 200,000 iterations. The loss function comprised three components. The first component is the standard Euclidean loss for the targets and the predictions (pixel loss; ℓ_p). The second component is the Euclidean loss for the

feature-transformed targets and the feature-transformed predictions (feature loss) (Johnson et al., 2016):

$$\ell_f = \frac{1}{n} \sum_{i,j,k} (\phi(t)_{i,j,k} - \phi(y)_{i,j,k})^2 \tag{5.1}$$

where n is the total number of features, $\phi(t)_{i,j,k}$ is a feature of the targets and $\phi(y)_{i,j,k}$ is a feature of the predictions. Similar to Johnson et al. (2016), we used the outputs of the fourth layer of a 16-layer DNN (relu_2_2 outputs of the VGG-16 pretrained model) (Simonyan & Zisserman, 2014) to feature transform the targets and the predictions. The third component is the total variation loss for the predictions:

$$\ell_{tv} = \sum_{i,j} ((y_{i+1,j} - y_{i,j})^2 + (y_{i,j+1} - y_{i,j})^2)^{0.5} \tag{5.2}$$

where $y_{i,j}$ is a pixel of the predictions. A weighted combination of these components resulted in the following loss function:

$$\ell = \lambda_p \ell_p + \lambda_f \ell_f + \lambda_{tv} \ell_{tv} \tag{5.3}$$

where we set $\lambda_p = \lambda_f = 1$ and $\lambda_{tv} = 0.00001$.

The use of the feature loss to train models for image transformation tasks was recently proposed by Johnson et al. (2016). In the context of super-resolution, Johnson et al. (2016) found that replacing pixel loss with feature loss gives visually pleasing results

at the expanse of image quality because of the artefacts introduced by the feature loss.

In the context of sketch inversion, our preliminary experiments showed that combining feature loss and pixel loss increases image quality while maintaining visual pleasantness. Furthermore, we observed that a small amount of total variation loss further removes the artefacts that are introduced by the feature loss. Therefore, we used the combination of the three losses in the final experiments. The quantitative results of the preliminary experiments in which the models were trained by using only the feature loss are provided in Supplementary Materials.

5.3 Results

Validation

First, we qualitatively tested the models by visual inspection of the synthesized face images (Figure 5.2). Synthesized face images matched the ground truth photographs closely and persons in the images were easily recognizable in most cases. Among the three styles of sketch models, the line sketch model (Figure 5.2, first column) captured the highest level of detail in terms of the face structure, whereas the synthesized inverse sketches of the color sketch model (Figure 5.2, third column) had less structural detail but was able to better reproduce the color information in the ground truth images compared to the inverted sketches of the line sketch model. Sketches synthesized by the grayscale model (Figure 5.2, second column) were less detailed than those synthesized by the line sketch model. Furthermore, the color content was less accurate in sketches synthesized by the grayscale model than those synthesized by both the color sketch and the line sketch models. We found that the line model performed impressively in terms of matching the hair and skin color of the

individuals even when the line sketches did not contain any color information. This may indicate that along with taking advantage of the luminance differences in the sketches to infer coloring, the model was able to learn color properties often associated with high-level face features of different ethnicities.

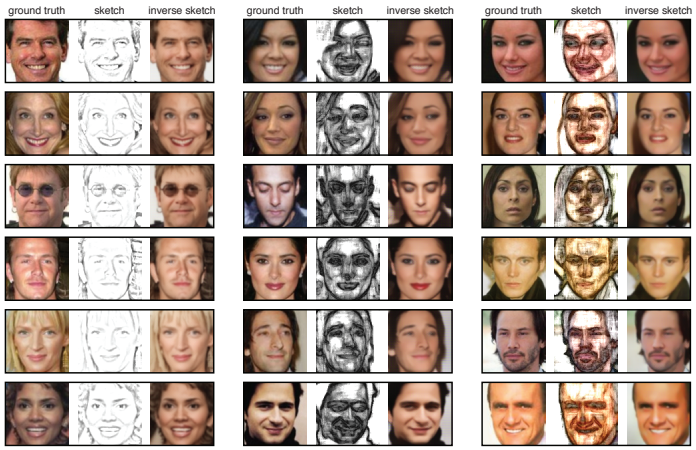

Figure 5.2: Examples of the synthesized inverse sketches from the LFW dataset. Each distinct column shows examples from different sketch styles models, i.e. line sketch model (column 1), grayscale sketch model (column 2) and colour sketch model (column 3). First image in each column is the ground truth, the second image is the generated sketch and the third one is the synthesized inverse sketch.

Then, we quantitatively tested the models by comparison of the peak signal to noise ratio (PSNR), structural similarity (SSIM) and standard Pearson product-moment correlation coefficient R of the synthesized face images (Wang, Bovik, Sheikh & Simoncelli, 2004) (Table 5.2). PSNR measures the physical quality of an image. It is defined as the ratio between the peak power of the image and the power of the noise in the image (Euclidean distance between the image and the reference image):

$$\text{PSNR} = \frac{1}{3}\sum_{k} 10\log_{10}\frac{\max \text{DR}^2}{\frac{1}{m}\sum_{i,j}(t_{i,j,k} - y_{i,j,k})^2} \tag{5.4}$$

where DR is the dynamic range, and m is the total number of pixels in each of the three color channels. SSIM measures the perceptual quality of an image. It is defined as the multiplicative combination of the similarities between the image and the reference image in terms of contrast, luminance and structure:

$$\text{SSIM} = \frac{1}{3}\sum_{k}\frac{1}{m}\sum_{i,j}\frac{(2\mu(t_{i,j,k})\mu(y_{i,j,k}) + C_1)(2\sigma(t_{i,j,k}, y_{i,j,k})C_2)}{(\mu(t_{i,j,k})^2\mu(y_{i,j,k})^2 + C_1)(2\sigma(t_{i,j,k})^2\sigma(y_{i,j,k})^2C_2)} \tag{5.5}$$

where $\mu(t_{i,j,k})$, $\mu(y_{i,j,k})$, $\sigma(t_{i,j,k})$, $\sigma(y_{i,j,k})$ and $\sigma(t_{i,j,k}, y_{i,j,k})$ are means, standard deviations and cross-covariances of windows centered around i and j. Furthermore, $C_1 = (0.01 \max \text{DR})^2$ and $C_2 = (0.03 \max \text{DR})^2$. Quality of a dataset is defined as the mean quality over the images in the dataset.

The inversion of the line sketches resulted in the highest quality face images for all three measures (20.12 for PSNR, 0.86 for SSIM and 0.93 for R). In contrast the inversion of the grayscale sketches resulted in the lowest quality face images for all measures (17.65 for PSNR, 0.65 for SSIM and 0.75 for R). This shows that both the physical and the perceptual quality of the inverted sketch images produced by the line sketch network was superior than those by the other sketch styles.

Finally, we tested how well the line sketch inversion model can be transferred to the task of synthesizing face images from sketches that are hand-drawn and not generated using the same methods

Table 5.2: Comparison of physical (PSNR), perceptual (SSIM) and correlational (R) quality measures for the inverse sketches synthesized by the line, grayscale and color sketch-style models. $x \pm m$ shows the mean $\pm$ the bootstrap estimate of the standard error of the mean.

	PSNR	SSIM	R
Line	**20.1158 ± 0.0231**	**0.8583 ± 0.0003**	**0.9298 ± 0.0005**
Grayscale	17.6567 ± 0.0263	0.6529 ± 0.0008	0.7458 ± 0.0020
Color	19.2029 ± 0.0293	0.7154 ± 0.0008	0.8087 ± 0.0017

that were used to train the model. We considered only the line sketch model since the contents of the hand-drawn sketch database that we used (Wang & Tang, 2009) were most similar to the line sketches.

We found that the line sketch inversion model can solve this inductive transfer task almost as good as it can solve the task that it was trained on (Figure 5.3). Once again, the model synthesized photorealistic face images. While color was not always synthesized accurately, other elements such as form, shape, line, space and texture were often synthesized well. Furthermore hair texture and style, which posed a problem in most previous studies, was very well handled by our CSI model. We observed that the dark-edged pencil strokes in the hand-drawn sketches that were not accompanied by shading resulted in less realistic inversions (compare e.g nose areas of sketches in the first and second rows with those in the third row in Figure 5.3). This can be explained by the lack of such features in the training data of the line sketch model, and can be easily overcome by including training examples more closely resembling the drawing style of the sketch artists.

For all the samples from the CUFS database, the PSNR, the SSIM index and the R of the synthesized face images were 13.42, 0.52, and 0.67, respectively (Table 5.3). Among the three sub-databases of the CUFS database, the quality of the synthesized images from the CUHK dataset was the highest in terms of the PSNR (15.07)

Figure 5.3: Examples of the synthesized inverse sketches from the CUFS database. First image in each column is the ground truth, the second image is the sketch hand-drawn by an artist and the third one is the inverse sketch that was synthesized by the line sketch model.

Table 5.3: Comparison of physical (PSNR), perceptual (SSIM) and correlational (R) quality measures for the inverse sketches synthesized from the sketches in the CUFS database and its subdatabases. $x \pm m$ shows the mean $\pm$ the bootstrap estimate of the standard error of the mean.

	PSNR	**SSIM**	**R**
CUHK (6)	**15.0675 ± 0.3958**	0.5658 ± 0.0099	**0.8264 ± 0.0269**
AR (6)	13.8687 ± 0.7009	**0.5684 ± 0.0277**	0.7667 ± 0.0314
XM2GTS (6)	11.3293 ± 1.2156	0.4231 ± 0.0272	0.4138± 0.1130
All (18)	13.4218 ± 0.6123	0.5191 ± 0.0207	0.6690 ± 0.0591

and R (0.83). While the PSNR and R values for the AR dataset was lower than those of the CUHK dataset, SSIM did not differ between the two datasets. The lowest quality inverted sketches were produced from the sample sketches of the XM2GTS database (with 13.42 for PSNR, 0.42 for SSIM and 0.41 for R).

Figure 5.4: Self-portrait sketches and synthesized inverse sketches along with a reference painting or photograph of famous Dutch artists: Rembrandt (top), Vincent van Gogh (middle) and M. C. Escher (bottom). Sketches: i) Self-Portrait with Beret, Wide-Eyed by Rembrandt, 1630, etching. ii) Two Self-portraits and Several Details by Vincent van Gogh, 1886, pencil on paper. iii) Self-Portrait by M.C. Escher, 1929, lithograph on gray paper. Reference paintings: i) Self-Portrait by Rembrandt, 1630, oil painting on copper. ii) Self-Portrait with Straw Hat by Vincent van Gogh, 1887, oil painting on canvas.

Applications

Fine Arts

In many cases self-portrait studies allow us a glimpse of what famous artists looked like through the artists' own perspective. Since there are no photographic records of many artists (in particular of those who lived before the 19th century during which the photography was invented and became widespread) self-portrait sketches and paintings are the only visual records that we have of many artists. Converting the sketches of the artists into photographs using a DNN that was trained on tens of thousands

of face sketch-photograph pairs results in very interesting end-products.

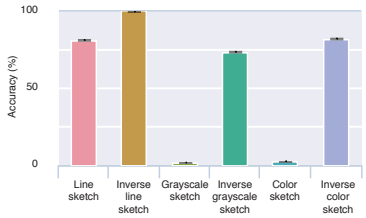

Figure 5.5: Identification accuracies for line, grayscale and color sketches, and for inverse sketches synthesized by the corresponding models. Error bars show the bootstrap estimates of the standard errors.

Here we used our DNN-based approach to synthesize photographs of famous Dutch artists Rembrandt, Vincent van Gogh and M. C. Escher from their self-portrait sketches (Figure 5.4). To the best of our knowledge, the synthesized photorealistic images of these artists are the first of their kind.

Our qualitative assesment revealed that, the inverted sketch of Rembrandt synthesized from his 1630 sketch indeed resembles himself in his paintings (particulary his self-portrait painting from 1630), and Escher's to his photographs. We found that the inverted sketch of van Gogh synthesized from his 1886 sketch was the most realistic synthesized photograph among those of the three artists, albeit not closely matching his self-portrait paintings of a distinct post-impressionist style.

Although we do not have a quantitative way to measure the accuracy of the results in this case, results demonstrate that the artistic style of the input sketches influence the quality of the produced photorealistic images. Generating new training sketch data to match more closely to the sketch style of a specific artist of interest (e.g. by using the method proposed by Zhang et al.

For simplicity, although different methods were used to produce these artworks, we refer to them as sketches.

(2016)), and training the network with these sketches would overcome this limitation.

Sketching is one of the most important training methods that artist use to develop their skills. Converting sketches into photorealistic images would allow the artists in training to see and evaluate the accuracy of their sketches clearly and easily which can in turn become an efficient training tool. Furthermore, sketching is often much faster than producing a painting. When for example the sketch is based on imagination rather than a photograph, deep sketch inversion can provide a photorealistic guideline (or even an end-product, if digital art is being produced) and can speed up the production process of artists. Figure 5.3, which shows the inverted sketches by contemporary artists that produced the sketches in the CUFS database, further demonstrates this type of application. The current method can be developed into a smartphone/tablet or computer application for common use.

Forensic Arts

In cases where no other representation of a suspect exists, sketches drawn by forensic artists based on eye-witness accounts are frequently used by the law enforcement. However, direct use of sketches for automatically identifying suspects from databases containing photographs does not work well because these two face representations are too different to allow a direct comparison (Wang, Tao, Gao, Li & Li, 2013a). Inverting a sketch to a photograph makes this task much easier by reducing the difference between these two alternative representations, enabling a direct automatized comparison (Wang & Tang, 2009).

To evaluate the potential use of our system for forensic applications, we performed an identification analysis (Figure 5.5). In this analysis, we evaluated the accuracy of identifying a target face image in a very large set of candidate face images (LFW dataset

containing over 11,000 images) from an (inverse) face sketch. The identification accuracies for the synthesized faces were always significantly higher than those for the corresponding sketched faces ($p << 0.05$, binomial test). While the identification accuracies for the color and grayscale sketches were very low (2.38% and 1.42%, respectively), those for the synthesized color and grayscale inverse sketches were relatively high (82.29% and 73.81%, respectively). On the other hand, identification accuracy of line sketches was already high, at 81.14% before inversion. Synthesizing inverse sketches from line sketches raised the identification accuracy to an almost perfect level (99.79%).

5.4 Conclusion

In this study we developed sketch datasets, complementing well known unconstrained benchmarking datasets (Liu et al., 2015; Learned-Miller et al., 2016), developed DNN models that can synthesize face images from sketches with state-of-the-art performance and proposed applications of our CSI model in fine arts, art history and forensics. We foresee further computer vision applications of the developed methods for non-face images and various other sketch-like representations, as well as cognitive neuroscience applications for the study of cognitive phenomena such as perceptual filling in (Vergeer, Anstis & van Lier, 2015; Anstis, Vergeer & Lier, 2012) and the neural representation of complex stimuli (Güçlü & van Gerven, 2015, 2017a).

Supplementary Materials

Table 5.4: Comparison of physical (PSNR), perceptual (SSIM) and correlational (R) quality measures for the inverse sketches synthesized by the line, grayscale and color sketch-style models trained using feature loss alone. $x \pm m$ shows the mean $\pm$ the bootstrap estimate of the standard error of the mean.

	PSNR	**SSIM**	**R**
Line	14.8956 ± 0.0207	0.5931 ± 0.0006	0.6023 ± 0.0017
Grayscale	17.1654 ± 0.0277	0.6301 ± 0.0008	0.7175 ± 0.0022
Color	$\mathbf{18.9884 \pm 0.0296}$	$\mathbf{0.7072 \pm 0.0008}$	$\mathbf{0.7976 \pm 0.0019}$

Table 5.5: Comparison of physical (PSNR), perceptual (SSIM) and correlational (R) quality measures for the inverse sketches synthesized from the sketches in the CUFS database and its subdatabases with the line sketch model trained using feature loss alone. $x \pm m$ shows the mean $\pm$ the bootstrap estimate of the standard error of the mean.

	PSNR	**SSIM**	**R**
CUHK (6)	$\mathbf{14.6213 \pm 0.4061}$	0.5358 ± 0.0216	$\mathbf{0.8295 \pm 0.0200}$
AR (6)	14.1721 ± 0.4127	$\mathbf{0.5608 \pm 0.0232}$	0.7811 ± 0.0217
XM2GTS (6)	11.7158 ± 1.3050	0.4096 ± 0.0258	0.3817 ± 0.1341
All (18)	13.5030 ± 0.5639	0.5021 ± 0.0205	0.6641 ± 0.0658

6

Reconstructing perceived faces from brain activations with deep adversarial neural decoding

This chapter is based on Güçlütürk, Y., Güçlü, U., Seeliger, K., Bosch, S., van Lier, R., and van Gerven, M. (2017). Reconstructing perceived faces from brain activations with deep adversarial neural decoding. In Neural Information Processing Systems. https://papers.nips.cc/paper/7012-reconstructing-perceived-faces-from-brain-activations-with-deep-adversarial-neural-decoding.pdf

6.1 Introduction

A key objective in sensory neuroscience is to characterize the relationship between perceived stimuli and brain responses. This relationship can be studied with neural encoding and neural decoding in functional magnetic resonance imaging (fMRI) (Naselaris et al., 2011). The goal of neural encoding is to predict brain responses to perceived stimuli (van Gerven, 2017). Conversely, the goal of neural decoding is to classify (Haxby, 2001; Kamitani & Tong, 2005), identify (Mitchell et al., 2008; Kay et al., 2008) or reconstruct (Thirion et al., 2006; Miyawaki et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011; Cowen et al., 2014) perceived stimuli from brain responses.

The recent integration of deep learning into neural encoding has been a very successful endeavor (Yamins & DiCarlo, 2016; Kriegeskorte, 2015). To date, the most accurate predictions of brain responses to perceived stimuli have been achieved with convolutional neural networks (Yamins et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014; Güçlü & van Gerven, 2015; Cichy, Khosla, Pantazis, Torralba & Oliva, 2016; Güçlü et al., 2016; Güçlü & van Gerven, 2017b; Eickenberg, Gramfort, Varoquaux & Thirion, 2017), leading to novel insights about the functional organization of neural representations. At the same time, the use of deep learning as the basis for neural decoding has received less widespread attention. Deep neural networks have been used for classifying or identifying stimuli via the use of a deep encoding model (Güçlü & van Gerven, 2015, 2017a) or by predicting intermediate stimulus features (Horikawa & Kamitani, 2017b, 2017a). Deep belief networks and convolutional neural networks have been used to reconstruct basic stimuli (handwritten characters and geometric figures) from patterns of brain activity (van Gerven et al., 2010; Du, Du & He, 2017). To date, going beyond such mostly retinotopy-driven reconstructions and reconstructing

complex naturalistic stimuli with high accuracy have proven to be difficult.

The integration of deep learning into neural decoding is an exciting approach for solving the reconstruction problem, which is defined as the inversion of the (non)linear transformation from perceived stimuli to brain responses to obtain a reconstruction of the original stimulus from patterns of brain activity alone. Reconstruction can be formulated as an inference problem, which can be solved by maximum a posteriori estimation. Multiple variants of this formulation have been proposed in the literature (Thirion et al., 2006; Naselaris et al., 2009; Güçlü & van Gerven, 2013; Schoenmakers et al., 2013; Schoenmakers, Güçlü, van Gerven & Heskes, 2015). At the same time, significant improvements are to be expected from deep neural decoding given the success of deep learning in solving image reconstruction problems in computer vision such as colorization (Zhang et al., 2016), face hallucination (Güçlütürk, Güçlü, van Lier & van Gerven, 2016a), inpainting (Pathak, Krähenbühl, Donahue, Darrell & Efros, 2016) and super-resolution (Ledig et al., 2016).

Here, we present a new approach by combining probabilistic inference with deep learning, which we refer to as deep adversarial neural decoding (DAND). Our approach first inverts the linear transformation from latent features to observed responses with maximum a posteriori estimation. Next, it inverts the nonlinear transformation from perceived stimuli to latent features with adversarial training and convolutional neural networks. An illustration of our model is provided in Figure 6.1. We show that our approach achieves state-of-the-art reconstructions of perceived faces from the human brain.

6.2 Materials and Methods

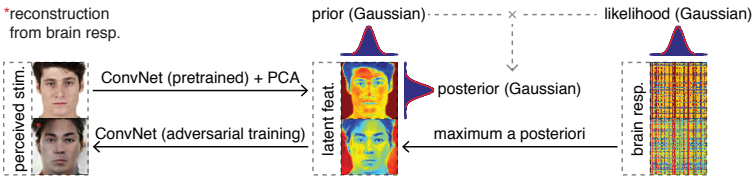

Figure 6.1: An illustration of our approach to solve the problem of reconstructing perceived stimuli from brain responses by combining probabilistic inference with deep learning.

Problem Statement

Let $\mathbf{x} \in \mathbb{R}^{h \times w \times c}$, $\mathbf{z} \in \mathbb{R}^{p}$, $\mathbf{y} \in \mathbb{R}^{q}$ be a stimulus, feature, response triplet, and $\phi : \mathbb{R}^{h \times w \times c} \rightarrow \mathbb{R}^{p}$ be a latent feature model such that $\mathbf{z} = \phi(\mathbf{x})$ and $\mathbf{x} = \phi^{-1}(\mathbf{z})$. Without loss of generality, we assume that all of the variables are normalized to have zero mean and unit variance.

We are interested in solving the problem of reconstructing perceived stimuli from brain responses:

$$\widehat{\mathbf{x}} = \phi^{-1}(\arg\max_{\mathbf{z}} \Pr(\mathbf{z}|\mathbf{y})) \tag{6.1}$$

where $\Pr(\mathbf{z}|\mathbf{y})$ is the posterior. We reformulate the posterior through Bayes' theorem:

$$\widehat{\mathbf{x}} = \phi^{-1}(\arg\max_{\mathbf{z}} [\Pr(\mathbf{y}|\mathbf{z}) \Pr(\mathbf{z})]) \tag{6.2}$$

where $\Pr(\mathbf{y}|\mathbf{z})$ is the likelihood, and $\Pr(\mathbf{z})$ is the prior. In the following subsections, we define the latent feature model, the likelihood and the prior.

Latent Feature Model

We define the latent feature model $\phi(\mathbf{x})$ by modifying the VGG-Face pretrained model (Parkhi, Vedaldi & Zisserman, 2015). This model is a 16-layer convolutional neural network, which was trained for face recognition. First, we truncate it by retaining the first 14 layers and discarding the last two layers of the model. At this point, the truncated model outputs 4096-dimensional latent features. To reduce the dimensionality of the latent features, we then combine the model with principal component analysis by estimating the loadings that project the 4096-dimensional latent features to the first 699 principal component scores (maximum number of components given the number of training observations) and adding them at the end of the truncated model as a new fully-connected layer. At this point, the combined model outputs 699-dimensional latent features.

Following the ideas presented in (Goodfellow et al., 2014; Radford, Metz & Chintala, 2015; Dosovitskiy & Brox, 2016), we define the inverse of the feature model $\phi^{-1}(\mathbf{z})$ (i.e., the image generator) as a convolutional neural network which transforms the 699-dimensional latent variables to $64 \times 64 \times 3$ images and estimate its parameters via an adversarial process. The generator comprises five deconvolution layers: The ith layer has 2^{10-i} kernels with a size of 4×4, a stride of 2×2, a padding of 1×1, batch normalization and rectified linear units. Exceptions are the first layer which has a stride of 1×1, and no padding; and the last layer which has three kernels, no batch normalization (Ioffe & Szegedy, 2015) and hyperbolic tangent units. Note that we do use the inverse of the loadings in the generator.

To enable adversarial training, we define a discriminator (ψ) along with the generator. The discriminator comprises five convolution layers. The ith layer has 2^{5+i} kernels with a size of 4×4, a stride of 2×2, a padding of 1×1, batch normalization and leaky rectified linear units with a slope of 0.2 except for the first layer which

has no batch normalization and last layer which has one kernel, a stride of 1×1, no padding, no batch normalization and a sigmoid unit.

We train the generator and the discriminator by pitting them against each other in a two-player zero-sum game, where the goal of the discriminator is to discriminate stimuli from reconstructions and the goal of the generator is to generate reconstructions that are indiscriminable from original stimuli. This ensures that reconstructed stimuli are similar to target stimuli on a pixel level and a feature level.

The discriminator is trained by iteratively minimizing the following discriminator loss function:

$$L_{\text{dis}} = -\mathbb{E}[\log(\psi(\mathbf{x})) + \log(1 - \psi(\phi^{-1}(\mathbf{z})))] \tag{6.3}$$

where ψ is the output of the discriminator which gives the probability that its input is an original stimulus and not a reconstructed stimulus. The generator is trained by iteratively minimizing a generator loss function, which is a linear combination of an adversarial loss function, a feature loss function and a stimulus loss function:

$$\begin{aligned} L_{\text{gen}} = &- \lambda_{\text{adv}} \underbrace{\mathbb{E}[\log(\psi(\phi^{-1}(\mathbf{z})))]}_{L_{\text{adv}}} + \lambda_{\text{fea}} \underbrace{\mathbb{E}[\|\xi(\mathbf{x}) - \xi(\phi^{-1}(\mathbf{z}))\|^2]}_{L_{\text{fea}}} + \\ &\lambda_{\text{sti}} \underbrace{\mathbb{E}[\|\mathbf{x} - \phi^{-1}(\mathbf{z})\|^2]}_{L_{\text{sti}}} \end{aligned} \tag{6.4}$$

where ξ is the relu3_3 outputs of the pretrained VGG-16 model (Simonyan & Zisserman, 2014; Johnson et al., 2016). Note that the targets and the reconstructions are lower resolution (i.e., 64×64) than the images that are used to obtain the latent features (i.e., 224×224).

Likelihood and Prior

We define the likelihood as a multivariate Gaussian distribution over $\mathbf{y}$:

$$\Pr(\mathbf{y}|\mathbf{z}) = \mathcal{N}_{\mathbf{y}}(\mathbf{B}^\top \mathbf{z}, \mathbf{\Sigma}) \tag{6.5}$$

where $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q) \in \mathbb{R}^{p \times q}$ and $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_q^2) \in \mathbb{R}^{q \times q}$. Here, the features $\times$ voxels matrix $\mathbf{B}$ contains the learnable parameters of the likelihood in its columns $\boldsymbol{\beta}_i$ (which can also be interpreted as regression coefficients of a linear regression model, which predicts $\mathbf{y}$ from $\mathbf{z}$).

We estimate the parameters with ordinary least squares, such that $\widehat{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}_i} \mathbb{E}[||y_i - \boldsymbol{\beta}_i^\top \mathbf{z}||^2]$ and $\widehat{\sigma}_i^2 = \mathbb{E}[||y_i - \widehat{\boldsymbol{\beta}}_i^\top \mathbf{z}||^2]$.

We define the prior as a zero mean and unit variance multivariate Gaussian distribution $\Pr(\mathbf{z}) = \mathcal{N}_{\mathbf{z}}(\mathbf{0}, \mathbf{I})$.

Posterior

To derive the posterior, we first reformulate the likelihood as a multivariate Gaussian distribution over $\mathbf{z}$. That is, after taking out constant terms with respect to $\mathbf{z}$ from the likelihood, it immediately becomes proportional to the canonical form Gaussian over $\mathbf{z}$

with $\nu = \mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{y}$ and $\Lambda = \mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{B}^{\top}$, which is equivalent to the standard form Gaussian with mean $\Lambda^{-1}\nu$ and covariance Λ^{-1}.

This allows us to write:

$$\Pr(\mathbf{z}|\mathbf{y}) \propto \mathcal{N}_{\mathbf{z}}\left(\Lambda^{-1}\nu, \Lambda^{-1}\right)\mathcal{N}_{\mathbf{z}}(\mathbf{0}, \mathbf{I}) \tag{6.6}$$

Next, recall that the product of two multivariate Gaussians can be formulated in terms of one multivariate Gaussian (Petersen & Pedersen, 2012). That is, $\mathcal{N}_{\mathbf{z}}(\mathbf{m}_1, \boldsymbol{\Sigma}_1)\mathcal{N}_{\mathbf{z}}(\mathbf{m}_2, \boldsymbol{\Sigma}_2) \propto \mathcal{N}_{\mathbf{z}}(\mathbf{m}_c, \boldsymbol{\Sigma}_c)$ with $\mathbf{m}_c = \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{m}_1 + \boldsymbol{\Sigma}_2^{-1}\mathbf{m}_2\right)$ and $\boldsymbol{\Sigma}_c = \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)^{-1}$. By plugging this formulation into Equation (6.6), we obtain

$$\Pr(\mathbf{z}|\mathbf{y}) \propto \mathcal{N}_{\mathbf{z}}(\mathbf{m}_c, \boldsymbol{\Sigma}_c) \tag{6.7}$$

with

$\mathbf{m}_c = (\mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{B}^{\top} + \mathbf{I})^{-1}\mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{y}$ and $\boldsymbol{\Sigma}_c = (\mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{B}^{\top} + \mathbf{I})^{-1}$.

Recall that we are interested in reconstructing stimuli from responses by generating reconstructions from the features that maximize the posterior. Notice that the (unnormalized) posterior is maximized at its mean $\mathbf{m}_c$ since this corresponds to the mode for a multivariate Gaussian distribution. Therefore, the solution of the problem of reconstructing stimuli from responses reduces to the following simple expression:

$$\hat{\mathbf{x}} = \phi^{-1}\left((\mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{B}^{\top} + \mathbf{I})^{-1}\mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{y}\right) \tag{6.8}$$

6.3 Results

Datasets

We used the following datasets in our experiments:

fMRI Dataset We collected a new fMRI dataset, which comprises face stimuli and associated blood-oxygen-level dependent (BOLD) responses. The stimuli used in the fMRI experiment were drawn from Ma, Correll and Wittenbrink (2015), Strohminger et al. (2015), Langner et al. (2010) and other online sources, and consisted of photographs of front-facing individuals with neutral expressions. We measured BOLD responses (TR = 1.4 s, voxel size $= 2 \times 2 \times 2$ mm^3, whole-brain coverage) of two healthy adult subjects (S1: 28-year old female; S2: 39-year old male) as they were fixating on a target (0.6×0.6 degree) (Thaler, Schütz, Goodale & Gegenfurtner, 2013) superimposed on the stimuli (15×15 degrees). Each face was presented at 5 Hz for 1.4 s and followed by a middle gray background presented for 2.8 s. In total, 700 faces were presented twice for the training set, and 48 faces were repeated 13 times for the test set. The test set was balanced in terms of gender and ethnicity (based on the norming data provided in the original datasets). The experiment was approved by the local ethics committee (CMO Regio Arnhem-Nijmegen) and the subjects provided written informed consent in accordance with the Declaration of Helsinki.

The stimuli were preprocessed as follows: Each image was cropped and resized to 224×224 pixels. This procedure was organized such that the distance between the top of the image and the vertical center of the eyes was 87 pixels, the distance between the vertical center of the eyes and the vertical center of the mouth was 75 pixels, the distance between the vertical center of the mouth and the bottom of the image was 61 pixels, and the horizontal center of the eyes and the mouth was at the horizontal center of the image.

The fMRI data were preprocessed as follows: Functional scans were realigned to the first functional scan and the mean functional scan, respectively. Realigned functional scans were slice time corrected. Anatomical scans were coregistered to the mean functional scan. Brains were extracted from the coregistered anatomical scans. Finally, stimulus-specific responses were deconvolved from the realigned and slice time corrected functional scans with a general linear model (Mumford, Turner, Ashby & Poldrack, 2012). Here, deconvolution refers to estimating regression coefficients ($\mathbf{y}$) of the following GLMs: $\mathbf{y}^* = \mathbf{X}\mathbf{y}$, where $\mathbf{y}^*$ is raw voxel responses, $\mathbf{X}$ is HRF-convolved design matrix (one regressor per stimulus indicating its presence), and $\mathbf{y}$ is deconvolved voxel responses such that $\mathbf{y}$ is a vector of size $m \times 1$ with m denoting the number of unique stimuli, and there is one $\mathbf{y}$ per voxel.

CelebA Dataset This dataset (Liu et al., 2015) comprises 202599 in-the-wild portraits of 10177 people, which were drawn from online sources. The portraits are annotated with 40 attributes and five landmarks. We preprocessed the portraits as we preprocessed the stimuli in our fMRI dataset.

Implementation Details

Our implementation makes use of Chainer and Cupy with CUDA and cuDNN (Tokui et al., 2015) except for the following: The VGG-16 and VGG-Face pretrained models were ported to Chainer from Caffe (Jia et al., 2014). Principal component analysis was implemented in scikit-learn (Pedregosa et al., 2011). fMRI preprocessing was implemented in SPM (Friston, Ashburner, Kiebel, Nichols & Penny, 2006). Brain extraction was implemented in FSL (Jenkinson, Beckmann, Behrens, Woolrich & Smith, 2012).

We trained the discriminator and the generator on the entire CelebA dataset by iteratively minimizing the discriminator loss

function and the generator loss function in sequence for 100 epochs with Adam (Kingma & Ba, 2014). Model parameters were initialized as follows: biases were set to zero, the scaling parameters were drawn from $\mathcal{N}(\mathbf{1}, 2 \cdot 10^{-2}\mathbf{I})$, the shifting parameters were set to zero and the weights were drawn from $\mathcal{N}(\mathbf{1}, 10^{-2}\mathbf{I})$ (Radford et al., 2015). We set the hyperparameters of the loss functions as follows: $\lambda_{\text{a}}dv = 10^2$, $\lambda_{\text{d}}is = 10^2$, $\lambda_{\text{f}}ea = 10^{-2}$ and $\lambda_{\text{s}}ti = 2 \cdot 10^{-6}$ (Dosovitskiy & Brox, 2016). We set the hyperparameters of the optimizer as follows: $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^8$ (Radford et al., 2015).

We estimated the parameters of the likelihood term on the training split of our fMRI dataset.

Evaluation Metrics

We evaluated our approach on the test split of our fMRI dataset with the following metrics: First, the feature similarity between the stimuli and their reconstructions, where the feature similarity is defined as the Euclidean similarity between the features, defined as the relu7 outputs of the VGG-Face pretrained model. Second, the Pearson correlation coefficient between the stimuli and their reconstructions. Third, the structural similarity between the stimuli and their reconstructions (Wang et al., 2004). All evaluation was done on a held-out set not used at any point during model estimation or training. The voxels used in the reconstructions were selected as follows: For each test trial, n voxels with smallest residuals (on training set) were selected. n itself was selected such that reconstruction accuracy of remaining test trials was highest.

Reconstruction

We first demonstrate our results by reconstructing the stimulus images in the test set using i) the latent features and ii) the brain responses. Figure 6.2 shows 4 representative examples of the test stimuli and their reconstructions. The first column of both panels show the original test stimuli. The second column of both panels show the reconstructions of these stimuli $\mathbf{x}$ from the latent features $\mathbf{z}$ obtained by $\phi(\mathbf{x})$. These can be considered as an upper limit for the reconstruction accuracy of the brain responses since they are the best possible reconstructions that we can expect to achieve with a perfect neural decoder that can exactly predict the latent features from brain responses. The third and fourth columns of the figure show reconstructions of brain responses to stimuli of Subject 1 and Subject 2, respectively.

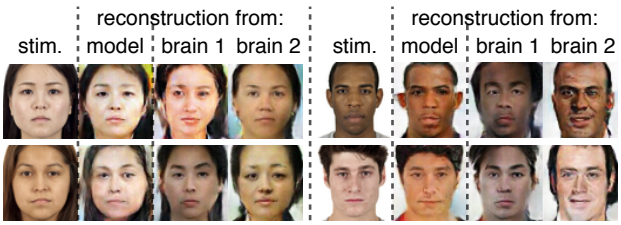

Figure 6.2: Reconstructions of the test stimuli from the latent features (model) and the brain responses of the two subjects (brain 1 and brain 2).

Visual inspection of the reconstructions from brain responses reveals that they match the test stimuli in several key aspects, such as gender, skin color and facial features. Table 6.1 shows the three reconstruction accuracy metrics for both subjects in terms of the ratio of the reconstruction accuracy from the latent features to the reconstruction accuracy from brain responses, which were significantly ($p < 0.05$, Student's t-test) above those for randomly sampled latent features (cf. 0.5181, 0.1532 and 0.5183, respectively).

Table 6.1: Reconstruction accuracy of the proposed decoding approach. The results are reported as the ratio of accuracy of reconstructing from latent features and brain responses.

	Feature similarity	Pearson correlation coefficient	Structural similarity
S1	0.6546 ± 0.0220	0.6512 ± 0.0493	0.8365 ± 0.0239
S2	0.6465 ± 0.0222	0.6580 ± 0.0480	0.8325 ± 0.0229

Furthermore, besides reconstruction accuracy, we tested the identification performance within and between groups that shared similar features (those that share gender or ethnicity as defined by the norming data were assumed to share similar features). Identification accuracies (which ranged between 57% and 62%) were significantly above chance-level (which ranged between 3% and 8%) in all cases ($p \ll 0.05$, Student's t-test). Furthermore, we found no significant differences between the identification accuracies when a reconstruction was identified among a group sharing similar features versus among a group that did not share similar features ($p > 0.79$, Student's t-test) (cf. Goesaert and Op de Beeck (2013)).

Visualization, Interpolation and Sampling

In the second experiment, we analyzed the properties of the stimulus features predictive of brain activations to characterize neural representations of faces. We first investigated the model representations to better understand what kind of features drive responses of the model. We visualized the features explaining the highest variance by independently setting the values of the first few latent dimensions to vary between their minimum and maximum values and generating reconstructions from these representations (Figure 6.3). As a result, we found that many of the latent features were coding for interpretable high level information such as age, gender, etc. For example, the first feature in Figure 6.3 appears to code for gender, the second one appears to code for hair color

and complexion, the third one appears to code for age, and the fourth one appears to code for two different facial expressions.

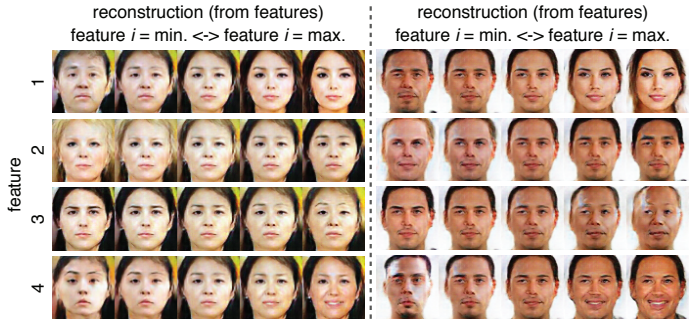

Figure 6.3: Reconstructions from features with single features set to vary between their minimum and maximum values.

We then explored the feature space that was learned by the latent feature model and the response space that was learned by the likelihood by systematically traversing the reconstructions obtained from different points in these spaces.

Figure 6.4*A* shows examples of reconstructions of stimuli from the latent features (rows one and four) and brain responses (rows two, three, five and six), as well as reconstructions from their interpolations between two points (columns three to nine). The reconstructions from the interpolations between two points show semantic changes with no sharp transitions.

Figure 6.4*B* shows reconstructions from latent features sampled from the model prior (first row) and from responses sampled from the response prior of each subject (second and third rows). The reconstructions from sampled representations are diverse and of high quality.

These results provide evidence that no memorization took place and the models learned relevant and interesting representations (Radford et al., 2015). Furthermore, these results suggest that neural

representations of faces might be embedded in a continuous and distributed space in the brain.

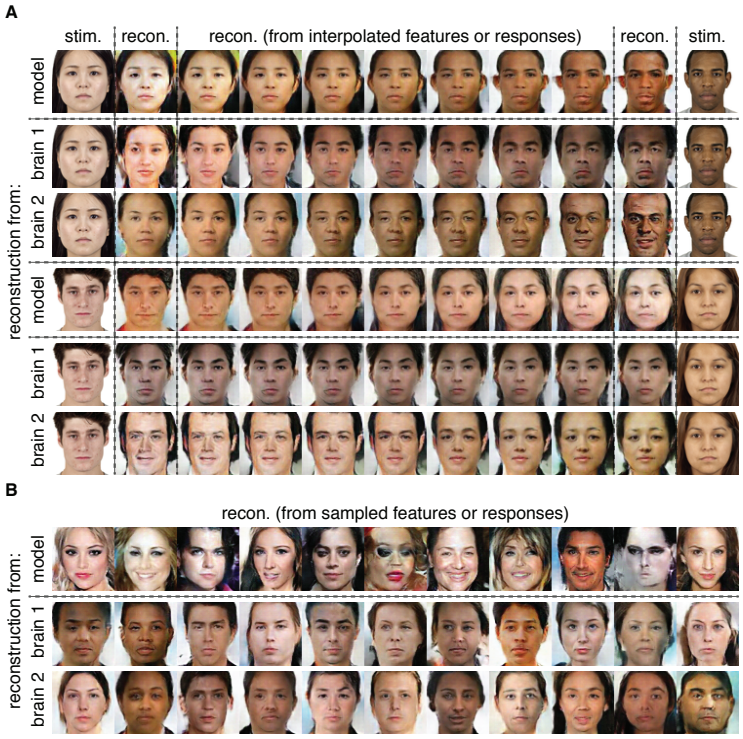

Figure 6.4: Reconstructions from interpolated (***A***) and sampled (***B***) latent features (model) and brain responses of the two subjects (brain 1 and brain 2).

Comparison Versus State-of-the-Art

In this section we qualitatively (Figure 6.5) and quantitatively (Table 6.2) compare the performance of our approach with two existing decoding approaches from the literature. Figure 6.5 shows example reconstructions from brain responses with three

We also experimented with the VGG-ImageNet pretrained model, which failed to match the reconstruction performance of the VGG-Face model, while their encoding performances were comparable in non-face related brain areas. We plan to further investigate other models in detail in future work.

different approaches, namely with our approach, the eigenface approach (Cowen et al., 2014; Lee & Kuhl, 2016) and the identity transform approach(van Gerven & Heskes, 2012; Schoenmakers et al., 2013). To achieve a fair comparison, the implementations of the three approaches only differed in terms of the feature models that were used, i.e. the eigenface approach had an eigenface (PCA) feature model and the identity transform approach had simply an identity transformation in place of the feature model.

Visual inspection of the reconstructions displayed in Figure 6.5 shows that DAND clearly outperforms the existing approaches. In particular, our reconstructions better capture the features of the stimuli such as gender, skin color and facial features. Furthermore, our reconstructions are more detailed, sharper, less noisy and more photorealistic than the eigenface and identity transform approaches. A quantitative comparison of the performance of the three approaches shows that the reconstruction accuracies achieved by our approach were significantly higher than those achieved by the existing approaches ($p \ll 0.05$, Student's t-test).

Figure 6.5: Reconstructions from the latent features brain responses of the two subjects (brain 1 and brain 2) using our decoding approach, as well as the eigenface and identity transform approaches for comparison.

Table 6.2: Reconstruction accuracies of the three decoding approaches. LF denotes reconstructions from latent features.

		Feature similarity	Pearson correlation coefficient	Structural similarity
Identity	S1	0.1254 ± 0.0031	0.4194 ± 0.0347	0.3744 ± 0.0083
	S2	0.1254 ± 0.0038	0.4299 ± 0.0350	0.3877 ± 0.0083
	LF	$\mathbf{1.0000 \pm 0.0000}$	$\mathbf{1.0000 \pm 0.0000}$	$\mathbf{1.0000 \pm 0.0000}$
Eigenface	S1	0.1475 ± 0.0043	0.3779 ± 0.0403	0.3735 ± 0.0102
	S2	0.1457 ± 0.0043	0.2241 ± 0.0435	0.3671 ± 0.0113
	LF	0.3841 ± 0.0149	0.9875 ± 0.0011	0.9234 ± 0.0040
DAND	S1	$\mathbf{0.1900 \pm 0.0052}$	$\mathbf{0.4679 \pm 0.0358}$	$\mathbf{0.4662 \pm 0.0126}$
	S2	$\mathbf{0.1867 \pm 0.0054}$	$\mathbf{0.4722 \pm 0.0344}$	$\mathbf{0.4676 \pm 0.0130}$
	LF	0.2895 ± 0.0137	0.7181 ± 0.0419	0.5595 ± 0.0181

Factors Contributing to Reconstruction Accuracy

Finally, we investigated the factors contributing to the quality of reconstructions from brain responses. All of the faces in the test set had been annotated with 30 objective physical measures (such as nose width, face length, etc.) and 14 subjective measures (such as attractiveness, gender, ethnicity, etc.). Among these measures, we identified five subjective measures that are important for face perception (Hahn & Perrett, 2014; Perrett, May & Yoshikawa, 1994; Birkás, Dzhelyova, Lábadi, Bereczkei & Perrett, 2014; Strom, Zebrowitz, Zhang, Bronstad & Lee, 2012; Little, Jones, Feinberg & Perrett, 2013; de Lurdes Carrito et al., 2016) as measures of interest and supplemented them with an additional measure of stimulus complexity. Complexity was included because of its important role in visual perception (Güçlütürk et al., 2016b). The selected measures were attractiveness, complexity, ethnicity, femininity, masculinity and prototypicality. Note that the complexity measure was not part of the dataset annotations and was defined as the Kolmogorov complexity of the stimuli, which was taken to be their compressed file sizes (Donderi & McFadden, 2005).

To this end, we correlated the reconstruction accuracies of the 48 stimuli in the test set (for both subjects) with their corresponding

measures (except for ethnicity) and used a two-tailed Student's *t*-test to test if the multiple comparison corrected (Bonferroni correction) *p*-value was less than the critical value of 0.05. In the case of ethnicity we used one-way analysis of variance to compare the reconstruction accuracies of faces with different ethnicities.

We were able to reject the null hypothesis for the measures complexity, femininity and masculinity, but failed to do so for attractiveness, ethnicity and prototypicality. Specifically, we observed a significant negative correlation (r = -0.3067) between stimulus complexity and reconstruction accuracy. Furthermore, we found that masculinity and reconstruction accuracy were significantly positively correlated (r = 0.3841). Complementing this result, we found a negative correlation (r = -0.3961) between femininity and reconstruction accuracy. We found no effect of attractiveness, ethnicity and prototypicality on the quality of reconstructions. We then compared the complexity levels of the images of each gender and found that female face images were significantly more complex than male face images ($p < 0.05$, Student's t-test), pointing to complexity as the factor underlying the relationship between reconstruction accuracy and gender. This result demonstrates the importance of taking stimulus complexity into account while making inferences about factors driving the reconstructions from brain responses.

6.4 Conclusion

In this study we combined probabilistic inference with deep learning to derive a novel deep neural decoding approach. We tested our approach by reconstructing face stimuli from BOLD responses at an unprecedented level of accuracy and detail, matching the target stimuli in several key aspects such as gender, skin color and facial features as well as identifying perceptual factors contributing to the reconstruction accuracy. Deep decoding approaches

such as the one developed here are expected to play an important role in the development of new neuroprosthetic devices that operate by reading subjective information from the human brain.

7 Summary and Discussion

7.1 Summary

Chapter 2 In Chapter 2, we investigated individual differences in complexity preferences of people based on their liking ratings for abstract digital art. The most prominent assumption regarding the relationship between complexity and liking has been an inverted U-curve relationship (Berlyne, 1971). Although the literature in art perception agrees that complexity is a critical stimulus property, recent studies have not been in a consensus about the direction of the influence of complexity in aesthetic appreciation of art (Nadal et al., 2010). To identify the causes of this disagreement, studying individual differences in complexity preferences is important.

With this goal, we conducted an experiment where we asked our participants to give complexity and liking ratings to a set of 144 grayscale digital abstract art images that varied in their complexity levels. Our participants were 19 women and 11 men, whose ages were between 20 and 32 years old (M: 21.3, SD: 3.5). The stimulus images were fractal-like statistical geometric patterns made up of circles, hexagons, squares and triangles, and were generated specifically for this study using a variant of the algorithm developed by Shier (2011).

We first calculated the average complexity preference of our participants. When plotted against complexity, these average liking ratings had an inverted U-curve shape. To identify different groups of participants who share complexity preferences, we then grouped people based on their liking ratings using k-means clustering algorithm. We then determined that the best grouping of participants was obtained when they were separated into two groups based on the analysis of the average silhouette values of the clusters. When we examined the average complexity ratings of people that were assigned to each cluster, we determined that one group had an average preference for high complexity images, whereas the other group had an average preference for

low-complexity images. Using generalized mixed models, we then tested whether a quadratic model, or a combination of two linear models (one for each cluster) explained the complexity-liking relationship, and found that the cluster-based model explained our data better. Furthermore, as a result of post-hoc analysis, we found a significant difference between the average liking rating reaction times of the two groups, where the group that preferred low-complexity images were faster. Additionally, we observed that the group who preferred low-complexity images were slightly older than the group that preferred high-complexity images. This difference was only marginally significant.

While providing us important insights into the role of complexity in visual preferences, the results that were found in Chapter 2 also emphasized the significance of selecting the correct analysis approach for the data at hand. Particularly in the field of empirical aesthetics, where subjective evaluations are at the focus of interest, this is an even more central issue, which should not be ignored.

In summary, the key outcomes of Chapter 2 were the following:

- We identified two groups of participants who had clearly opposite average complexity preferences for our highly controlled set of grayscale statistical geometric patterns.

- We showed that one of the best known rules of aesthetic preference, i.e. the inverted U-curve of preference for complexity, can in fact be an artifact that arises from selecting a non-ideal analysis method. By employing an averaging approach, most experimental aesthetics studies risk reaching conclusions about a non-existent average observer.

- Implications of this finding include a need for utilization of new analysis methods that take into account the differences between individuals, re-evaluation of established rules of

human aesthetic preferences and revision of existing theories.

Chapter 3 In Chapter 3, we extended the ideas that were developed in Chapter 2 by investigating individual differences in complexity preferences for music in a larger pool of participants. By doing so, we tested the generalizability of our findings to ecologically valid stimulus types and the auditory domain.

Here, we reanalyzed a data set that was part of a previous music preference study (Greenberg et al., 2015). It comprised liking ratings of 353 participants for 15-second-long 25 song excerpts belonging to 16 different music genres. Furthermore, several demographic and personality related measures of the participants were also available as part of this data set. Among these measures we included age, sex, EQ, SQ-R, brain type and AQ in our analyses.

Particularly, we first clustered participants into several groups using k-means clustering algorithm and identified the best grouping based on silhouette analysis. These analyses revealed that the best grouping of participants consisted of two clusters with opposite complexity preferences for music. We then compared the agreement between participants in their liking ratings before and after the clustering analysis by calculating intraclass correlation indices, and found that clustering resulted in groups of people that had higher agreement rates compared to the whole sample. Next, we repeated the same generalized mixed model analysis that we performed in Chapter 2, and found similar results, such that the cluster-based model explained the data better than the quadratic model. Following this, we performed a number of post-hoc analysis to identify distinguishing characteristics of the clusters. Our initial comparisons revealed that distribution of participants in terms of age, sex, brain type, and AQ significantly differed between the two clusters. Specifically, compared to the low complexity favoring group, the high complexity favoring group was

on average younger, had more males, more Type S people, and had a higher average AQ. Then applying a logistic regression analysis, we confirmed age and sex as significant factors, where males were identified to be 1.77 times more likely to have a preference for high-complexity music, and increasing age increased the likelihood of having a preference for low-complexity music.

In summary, the key outcomes of Chapter 3 were the following:

- We showed that the results regarding the duality of the complexity preferences that we obtained in Chapter 2 were not limited to the visual domain and highly controlled stimuli, but also were evident in the auditory domain with complex ecologically valid music stimuli.

- We identified age and gender as important factors in complexity preferences for music, as well as showing possible links between cognitive brain type, AQ, and complexity preferences for music.

- Once again, we showed that pooling preference data across participants may easily obscure relevant differential tendencies between groups of participants.

Chapter 4 Chapter 4 was the last chapter where we focused our investigations on the construction of complexity perception. In Chapter 4 we investigated the neural representations of complexity in the visual and auditory modalities. As we have seen in the Chapters 2 and 3, complexity has a substantial role in liking of both visual and auditory stimuli. Behavioral effects of complexity in task performance in several perceptual and cognitive tasks have been well studied. However, its neural representations in the sensory cortices are not well understood.

To investigate the neural representations of complexity in the sensory cortices of the human brain, we utilized encoding and

decoding analyses of two fMRI datasets in the visual (Experiment 1) and auditory (Experiment 2) modalities. The dataset of Experiment 1 was a preexisting dataset (Horikawa & Kamitani, 2017a), which comprises 1250 photographs from 200 object categories as visual stimuli and fMRI data of five healthy adult participants. For quantifying the complexity levels of the visual stimuli, three different computational complexity measures were used: mean maximum magnitude gradient, PNG compressed file size, self-similarity. For Experiment 2, we collected fMRI data from eight healthy adult participants while they listened to 144 29-second long song excerpts. To quantify the complexity levels of the music stimuli, three different computational complexity measures were used: FLAC compressed file size, Ogg Vorbis compressed file size, and event density. The data of both experiments were analyzed using the same methods. Encoding and decoding analyses were performed on the ROIs in the early and association sensory cortices of hyperaligned data of all participants in each experiment. In decoding analysis, we used ridge regression to predict complexity measures from stimulus-evoked responses of voxels in ROIs. In encoding analysis, we used linear regression to predict stimulus-evoked responses of voxels in ROIs from complexity measures. Statistical analyses were performed using permutation tests.

In Experiment 1, we found that visual complexity representations were present throughout the visual cortex. Sensitivity of ROIs in the visual cortex decreased in a gradual manner from lower to higher visual areas. Among the higher visual areas, decoding performance in PPA was the highest, followed by LOC and FFA. Furthermore, the overlap between voxels representing different complexity measures were higher in the early visual cortex compared to the associative visual cortex. We also observed that a large number of voxels in the higher visual areas showed increased activity as image complexity decreased. The results of Experiment 2 showed large parallels to those of Experiment 1. One difference was that the sensitivity difference between the lower and higher auditory cortices were more abrupt compared

to those in the visual areas. In the associative auditory cortex, we found that posterior STS, posterior STG, and TA2 regions had high decoding performances. We observed that decoding and encoding in anterior STS and STG regions were mostly close to chance level. Furthermore, we determined that representations of event density in music were less pronounced compared to those of Kolmogorov complexity measures throughout the auditory cortex.

In summary, the key outcomes of Chapter 4 were the following:

- We characterized the complexity representations throughout the auditory and visual sensory cortices.
- We quantified the complexity sensitivity of individual ROIs, demonstrating a change of sensitivity (from fine grained to coarser) in a gradient from lower to higher areas.
- We demonstrated that PPA had distributed representations of complexity comparable to or better than the ROIs in the early visual cortex, supporting the notion that global scene properties such as complexity plays an important role in scene processing.
- We showed that regions of the auditory cortex, which represent syntactic language complexity such as posterior STG/S, also represent music complexity.
- We showed that while early sensory cortices are responsive to all complexity measures, voxels in the ROIs of associative sensory cortices become more specialized along the sensory hierarchy.

Chapter 5 In Chapter 5 we moved on to a more application oriented approach and investigated reconstructions of stimuli. The problem that we tackled in this chapter was converting face sketches to photorealistic face images. Prior work in this field

has been mostly limited to highly controlled sketches with similar lighting conditions, frontal poses and neutral expressions. Here we suggested an approach that can handle variability in conditions such as lighting, pose, and facial expression.

In order to train the necessary models for this task, we first complemented two large face photograph datasets (Liu et al., 2015; Learned-Miller et al., 2016), which are often used for bench marking purposes in computer vision, with automatically generated sketch versions of each photograph. An important preprocessing step was alignment of the face images by using key points in faces such as the centers of the eyes and mouth. We generated three different sketch types: line, grayscale, and color sketch. We developed a model for each of these three sketch types. The architecture of all models were the same and comprised four convolutional layers, five residual blocks, two deconvolutional layers and another convolutional layer. A key aspect for training our models was the use of a weighted combination of loss functions including pixel loss, feature loss and total variation loss. We trained our models to minimize a weighted combination of these loss functions, which resulted in learning a mapping between input sketch images and their corresponding photographs. We then tested our reconstruction accuracy on images that were previously unseen by the model.

All models performed well in reconstructing faces, however reconstructions from the line sketch model looked the best. We tested how well our reconstructions would perform in an automatic recognition test and showed that synthesized images that were obtained by using the line sketch model had almost perfect recognition accuracy (99.79%). All reconstructions improved recognition accuracy compared to recognition using sketches only. Furthermore, besides testing our models with computer generated sketches, we also performed tests with hand-drawn sketch databases. While the quality of the reconstructions were high, they were relatively worse than the reconstructions of computer

generated sketches. Finally, we demonstrated photorealistic reconstructions of self-portrait sketches of Rembrandt, van Gogh, and Escher.

In summary, the key outcomes of Chapter 5 were the following:

- We developed DNN models that can synthesize face images from sketches with state-of-the-art performance.
- We proposed applications of our model in fine arts, art history and forensics.
- We reconstructed photorealistic portraits of famous Dutch artists to demonstrate the art historical and fine arts application prospects of our model.

In Chapter 6 we investigated another reconstruction application. In this chapter, instead of a simple to complex stimuli transformation as presented in the previous chapter, we focused on a neural decoding problem within the similar domain of face perception.

To address the problem of face decoding from human brain activity, we used several models and combined them in a probabilistic framework. For extracting face features, we used a pretrained deep neural network, which was trained for face recognition. Using another dataset, we developed and trained another deep neural network for transforming the extracted highly nonlinear face features back to face images. A final linear model was used for mapping face features that were extracted from the stimulus images to their corresponding voxel responses. Finally, we inverted this mapping using Bayes Rule, such that we could obtain face features for the given voxel responses. Then we used the model that could transform face features to face images to reconstruct the perceived face images from fMRI data of test subjects. To test our approach we collected an fMRI dataset comprising face stimuli and fMRI data of two healthy adult participants. Each

participant viewed 748 face images. Similar to the preprocessing of images in Chapter 5, all faces that were used in all models were aligned based on automatically identified facial landmarks.

We compared our reconstruction results to those of several control models, in which intermediate features were given by pixel values or eigenfaces. These comparisons revealed that our results were higher quality and more similar to the stimulus images both qualitatively and quantitatively. Furthermore, by visualizing the face features that were used to predict voxel responses, we observed that they were meaningful features representing properties such as age, facial expression, and ethnicity. Additionally, we explored the relation between reconstruction performance and several stimulus properties e.g. attractiveness, complexity, etc., and identified stimulus complexity as a significant factor contributing to the accuracy of reconstructions.

In summary, the key outcomes of Chapter 6 were the following:

- We presented a novel approach of reconstructing complex stimulus percepts by combining probabilistic inference with deep learning.

- We generate state-of-the-art reconstructions of perceived faces from brain activations.

- We showed that stimulus complexity was an important factor that influences decoding performance. This result demonstrates the importance of taking stimulus complexity into account while making inferences about factors driving the reconstructions from brain responses.

In this thesis, we aimed to investigate the effects of complexity on several aspects of human perception. While appreciation of stimulus and the role of complexity in art has been of special interest throughout our investigations (e.g. in Chapters 2, 3, 4

and 5), we also explored the importance of complexity in perception and reconstruction of percepts through the use of non-artistic stimuli (e.g. in Chapters 4 and 6). In Chapters 2 and 3, our main questions were: What is the influence of complexity on appreciation of visual art and music? Does this influence differ among different individuals? As a result, both for visual art and music, we identified two groups of participants who had clearly opposite average complexity preferences, and showed that the mid-level complexity was not the globally preferred complexity level. In Chapter 4, we asked the question: How is complexity represented across the sensory cortices? Answering this question, we characterized complexity representations throughout the auditory and visual cortices and revealed a change of complexity sensitivity from lower to higher sensory areas. In Chapter 5, we asked whether it would be possible to reconstruct a complex stimulus, i.e. a face photograph, from its simpler representation, i.e. a sketch, and found that this was possible with the help of sophisticated deep learning methods. Finally in Chapter 6, we developed a method for decoding faces from stimulus-evoked brain activations measured by fMRI and asked the question: Do differences in complexity levels of face stimuli influence the decoding performance of our method? We found that the answer was yes, and even among the same stimulus type, stimulus complexity was negatively correlated with the accuracy of reconstructions of perceived face images.

Outlook The objective of this thesis was to investigate the influence of complexity in construction and reconstruction of perceived stimuli. To this end, we asked several questions, and as expected, obtaining the answers to these questions generated some new research questions. In this section such questions will be discussed.

Results of Chapter 2 and Chapter 3, which revealed the duality of complexity preferences in vision and music, warrant the investigation of many interesting research questions. First one of such

questions concern the generalizability of the results to different stimulus types. Is the duality of complexity preferences limited to simple visual patterns and music of various genres or can it be observed with other sets of stimuli, for example within a specific genre of songs, or a varied set of paintings? To what extent do these results generalize? The examination of these question would require the use of different datasets in a similar fashion that we presented in Chapters 2 and 3. In fact, our results have already been replicated in the visual domain with grayscale patterns that have fractal properties as stimulus and western oil paintings (Bies, Blanc-Goldhammer, Boydston, Taylor & Sereno, 2016; Spehar et al., 2016; Hayn-Leichsenring et al., 2017), where each of these studies identified two or more clusters of participants with distinct preference patterns and related these results to measures related to stimulus complexity.

Another future research direction would deal with further characterizing the groups of participants with different preferences. For example, other related personality measures such as Big-Five personality traits (particularly openness as identified by previous research), need for cognition, and tolerance to ambiguity, and creativity can be explored with respect to their role in complexity preferences. Similarly, considering motivation levels, mood, and health conditions of participants, which have been shown as relevant factors in complexity perception in earlier studies, may help further understand these preferences.

Recently, Spehar et al. (2015) showed a link between visual sensitivity and visual preferences. This line of research could be extended for explaining the results of both Chapter 2 and 3. We can formulate the following research question in this direction: Does the preference in a modality also depend on sensory sensitivity, and if so, what type of sensitivity could potentially account for these differences? In the visual domain, sensitivity to amplitude spectra and spatial frequencies have already been identified (Spe-

har et al., 2015). In the auditory domain, we predict that pitch sensitivity could be relevant.

Another promising research question concerns the possibility of a supramodal mechanism of appreciation: Does the complexity preference in one modality also persist in the other modality, i.e. if a person likes complex music, would they also like complex visual art? If this is the case, this may suggest a different supramodal mechanism and an explanation other than sensory sensitivity. Such results would further necessitate moving towards neuroaesthetics theories encompassing different sensory modalities (Brattico et al., 2017; Marin, 2015).

In Chapter 4 we looked at the complexity representations in the sensory cortices. Given the demonstrated role of complexity in art perception, future studies should extend our investigations to whole brain analysis. Furthermore, the role of different complexity measures, such as subjective complexity in both visual and auditory domains or structural complexity in the visual domain could be investigated as well. Another interesting question regarding complexity representations in the brain involves exploring specific stimulus categories. Would the complexity representations (especially in the specialized visual cortices) remain the same when different categories of images are viewed? A similar question can be also posed for the auditory domain.

Photorealistic image synthesis methods that were presented in Chapter 5 have a rather large prospect of expansion to solve similar other problems. In fact there have already been several similar studies in the field (Sangkloy, Lu, Fang, Yu & Hays, 2016; Guo, Zhu, Xia, Wang & Liu, 2017). As we demonstrated in Chapter 6, deep learning methods similar to those that were used to reconstruct faces from sketches, can be successfully used in brain decoding studies. Furthermore, we foresee that commercial drawing applications that are based on the same or similar principles that we presented may become available in the near future.

In Chapter 6, we demonstrated the use of deep neural networks in combination with probabilistic inference for decoding faces from human brain with an unprecedented level of success. Using similar methods, we believe that it would be possible to decode even more complex perceived stimuli from the brain activity. An interesting research question combining Chapters 5 and 6 is: Would it be possible to decode faces from perceived face sketches? Previously, it has been shown that contours may result in the perception of nonexistent colors for example, in the case of the watercolor illusion (Pinna, Brelstaff & Spillmann, 2001) and after images (van Lier, Vergeer & Anstis, 2009). Although earlier visual regions have been implicated (Komatsu, 2006; Huang & Paradiso, 2008; van Lier et al., 2009), the underlying processes that lead to the perception of these absent colors have not been well established (Cornelissen, 2006; Hazenberg & van Lier, 2013). Therefore, it would be very interesting to see which brain regions play a role in the 'filling-in' of color information that lack in the face sketches, and how the processing and construction of color perception relates to the image reconstruction operations of the deep neural networks. In Chapter 6 we used fMRI data of the whole brain of the participants. However, we did not perform an analysis of the contribution of different brain regions to different aspects of the reconstructions. We plan to investigate neural representations of different face features (e.g. perceived age, sex, hair color, face structure, etc.) in upcoming studies.

7.2 Conclusion

It has been more than a century since the earliest studies on the relationship between complexity and perception. Throughout this time, there have been countless ideas about the effects of complexity on various perceptual phenomena, and rules that try to formalize these ideas. However, even today, one can witness the emergence of new results, and the refinement or the replace-

ment of existing rules. In particular, tackling this problem from different angles, such as computational, behavioral and neural, as we have embraced in this thesis, has been shown to be a promising approach. In the future, we expect the integration of such approaches to complexity to provide a more complete description thereof.

Bibliography

Ahissar, M. & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406. doi:10.1038/387401a0. (Cit. on p. 74)

Aks, D. J. & Sprott, J. C. (1996). Quantifying aesthetic preference for chaotic patterns. *Empirical Studies of the Arts*, *14*(1), 1–16. doi:10.2190/6v31-7m9r-t9l5-cdg9. (Cit. on pp. 15, 27)

Alpaugh, P. K. & Birren, J. E. (1977). Variables affecting creative contributions across the adult life span. *Human Development*, *20*(4), 240–248. doi:10.1159/000271559. (Cit. on pp. 16, 44, 70)

Anderson, N. S. & Leonard, J. A. (1958). The recognition, naming, and reconstruction of visual figures as a function of contour redundancy. *Journal of Experimental Psychology*, *56*(3), 262–270. doi:10.1037/h0044893. (Cit. on pp. 9, 74)

Anstis, S., Vergeer, M. & Lier, R. V. (2012). Looking at two paintings at once: Luminance edges can gate colors. *i-Perception*, *3*(8), 515–519. doi:10.1068/i0537sas. (Cit. on p. 128)

Arthur, D. & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In N. Bansal, K. Pruhs & C. Stein (Eds.), *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms, SODA 2007, new orleans, louisiana, usa, january 7-9, 2007* (pp. 1027–1035). SIAM. (Cit. on p. 32).

Augustin, M. D., Wagemans, J. & Carbon, C.-C. (2012a). All is beautiful? generality vs. specificity of word usage in visual aesthetics. *Acta Psychologica*, *139*(1), 187–201. doi:10.1016/j.actpsy.2011.10.004. (Cit. on p. 15)

Augustin, M. D., Carbon, C.-C. & Wagemans, J. (2012b). Artful terms: A study on aesthetic word usage for visual art versus film and music. *i-Perception*, *3*(5), 319–337. doi:10.1068/i0511aap. (Cit. on p. 15)

Bader, R. (2013). *Nonlinearities and synchronization in musical acoustics and music psychology*. Springer Berlin Heidelberg. doi:10.1007/978-3-642-36098-5. (Cit. on p. 104)

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N. & Wheelwright, S. (2003). The systemizing quotient: An investigation of adults with asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 361–374. doi:10.1098/rstb.2002.1206. (Cit. on p. 58)

Baron-Cohen, S. (2010). Empathizing, systemizing, and the extreme male brain theory of autism. In *Sex differences in the human brain, their underpinnings and implications* (pp. 167–175). Elsevier. doi:10.1016/b978-0-444-53630-3.00011-7. (Cit. on p. 59)

Baron-Cohen, S. & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. doi:10.1023/b:jadd.0000022607.19833.00. (Cit. on p. 58)

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. (2001). *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. doi:10.1023/a:1005653411471. (Cit. on p. 59)

Barron, F. (2012). *Creativity and psychological health: Origins of personal vitality and creative freedom*. Literary Licensing, LLC. (Cit. on p. 44).

Barron, F. & Welsh, G. S. (1952). Artistic perception as a possible factor in personality style: Its measurement by a figure preference test. *The Journal of Psychology*, *33*(2), 199–203. doi:10.1080/00223980.1952.9712830. (Cit. on p. 44)

Berlyne, D. E. (1971). *Aesthetics and psychobiology (the century psychology series)*. Appleton-Century-Crofts. (Cit. on pp. 5, 14, 26, 45, 54, 74, 75, 152).

Berlyne, D. E. (1963). Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *17*(3), 274–290. doi:10.1037/h0092883. (Cit. on pp. 5, 13, 26)

Berlyne, D. E. (1977). Psychological aesthetics, speculative and scientific. *Leonardo*, *10*(1), 56. doi:10.2307/1573634. (Cit. on p. 28)

Berlyne, D. E., Ogilvie, J. C. & Parham, L. C. (1968). The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *22*(5), 376–387. doi:10.1037/h0082777. (Cit. on pp. 5, 13, 26, 75)

Besson, M., Schön, D., Moreno, S., Santos, A. & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, *25*(3-4), 399–410. (Cit. on p. 10).

Beyeler, M. (2015). *Opencv with python blueprints*. Packt Publishing - ebooks Account. (Cit. on p. 117).

Bies, A. J., Blanc-Goldhammer, D. R., Boydston, C. R., Taylor, R. P. & Sereno, M. E. (2016). Aesthetic responses to exact fractals driven by physical complexity. *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2016.00210. (Cit. on p. 162)

Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T. & Perrett, D. I. (2014). Cross-cultural perception of trustworthiness: The effect of ethnicity features on evaluation of faces' observed trustworthiness across four samples. *Personality and Individual Differences*, *69*, 56–61. doi:10.1016/j.paid.2014.05.012. (Cit. on p. 147)

Birkhoff, G. D. (1933). *Aesthetic measure*. Harvard University Press. (Cit. on p. 45).

Bonneville-Roussy, A., Stillwell, D., Kosinski, M. & Rust, J. (2017). Age trends in musical preferences in adulthood: 1. conceptualization and empirical investigation. *Musicae Scientiae*, 102986491769157. doi:10.1177/1029864917691571. (Cit. on p. 70)

Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K. & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of Personality and Social Psychology*, *105*(4), 703–717. doi:10.1037/a0033770. (Cit. on p. 70)

Bosch, A., Zisserman, A. & Munoz, X. (2007). Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on image and video retrieval - CIVR '07*. ACM Press. doi:10.1145/1282280.1282340. (Cit. on pp. 13, 81)

Boselie, F. (1984). Complex and simple proportions and the aesthetic attractivity of visual patterns. *Perception*, *13*(2), 91–96. doi:10.1068/p130091. (Cit. on p. 45)

Boselie, F. & Leeuwenberg, E. (1985). Birkhoff revisited: Beauty as a function of effect and means. *The American Journal of Psychology*, *98*(1), 1. doi:10.2307/1422765. (Cit. on p. 13)

Brattico, P., Brattico, E. & Vuust, P. (2017). Global sensory qualities and aesthetic experience in music. *Frontiers in Neuroscience*, *11*. doi:10.3389/fnins.2017.00159. (Cit. on pp. 71, 163)

Braun, J., Amirshahi, S. A., Denzler, J. & Redies, C. (2013). Statistical image properties of print advertisements, visual artworks and images of architecture. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00808. (Cit. on pp. 13, 74, 75, 80)

Carballal, A., Santos, A., Romero, J., Machado, P., Correia, J. & Castro, L. (2016). Distinguishing paintings from photographs by complexity estimates. *Neural Computing and Applications*. doi:10.1007/s00521-016-2787-5. (Cit. on p. 13)

Carbon, C.-C. (2011). Cognitive mechanisms for explaining dynamics of aesthetic appreciation. *i-Perception*, *2*(7), 708–719. doi:10.1068/i0463aap. (Cit. on p. 14)

Carbon, C.-C. (2010). The cycle of preference: Long-term dynamics of aesthetic appreciation. *Acta Psychologica*, *134*(2), 233–244. doi:10.1016/j.actpsy.2010.02.004. (Cit. on p. 14)

Cataltepe, Z., Yaslan, Y. & Sonmez, A. (2007). Music genre classification using MIDI and audio features. *EURASIP Journal on Advances in Signal Processing*, *2007*(1), 036409. doi:10.1155/2007/36409. (Cit. on p. 12)

Chai, X. J. (2010). Scene complexity: Influence on perception, memory, and development in the medial temporal lobe. *Frontiers in Human Neuroscience*, *4*. doi:10.3389/fnhum.2010.00021. (Cit. on pp. 9, 10, 74)

Chaitin, G. J. (1975). Randomness and mathematical proof. *Scientific American*, *232*(5), 47–52. (Cit. on p. 2).

Chamberlain, R. & Wagemans, J. (2015). Visual arts training is linked to flexible attention to local and global levels of visual stimuli. *Acta Psychologica*, *161*, 185–197. doi:10.1016/j.actpsy.2015.08.012. (Cit. on p. 10)

Chamorro-Premuzic, T., Burke, C., Hsu, A. & Swami, V. (2010). Personality predictors of artistic preferences as a function of the emotional valence and perceived complexity of paintings. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(4), 196–204. doi:10.1037/a0019211. (Cit. on pp. 16, 27)

Chamorro-Premuzic, T., Reimers, S., Hsu, A. & Ahmetoglu, G. (2009). Who art thou? personality predictors of artistic preferences in a large UK sample: The importance of openness. *British Journal of Psychology*, *100*(3), 501–516. doi:10.1348/000712608x366867. (Cit. on pp. 27, 45)

Chatterjee, A. (2004). Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of Psychology and the Arts*, *4*(2), 55–60. (Cit. on p. 13).

Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B. & Bromberger, B. (2010). The assessment of art attributes. *Empirical Studies of the Arts*, *28*(2), 207–222. doi:10.2190/em.28.2.f. (Cit. on p. 11)

Cheng, Z., Yang, Q. & Sheng, B. (2015). Deep colorization. In *2015 IEEE international conference on computer vision (ICCV)*. IEEE. doi:10.1109/iccv.2015.55. (Cit. on p. 113)

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A. & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*(1). doi:10.1038/srep27755. (Cit. on p. 132)

Clatworthy, J., Buick, D., Hankins, M., Weinman, J. & Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. *British Journal of Health Psychology*, *10*(3), 329–358. doi:10.1348/135910705x25697. (Cit. on p. 28)

Cleridou, K. & Furnham, A. (2014). Personality correlates of aesthetic preferences for art, architecture, and music. *Empirical Studies of the Arts*, *32*(2), 231–255. doi:10.2190/em.32.2.f. (Cit. on pp. 16, 27, 45)

Corbit, J. D. & Garbary, D. J. (1995). Fractal dimension as a quantitative measure of complexity in plant development. *Proceedings of the Royal Society B: Biological Sciences*, *262*(1363), 1–6. doi:10.1098/rspb.1995.0168. (Cit. on p. 75)

Cornelissen, F. W. (2006). No functional magnetic resonance imaging evidence for brightness and color filling-in in early human visual cortex. *Journal of Neuroscience*, *26*(14), 3634–3641. doi:10.1523/jneurosci.4382-05.2006. (Cit. on p. 164)

Corrigall, K. A. & Schellenberg, E. G. (2015). Liking music: Genres, contextual factors, and individual differences. In *Art, aesthetics, and the brain* (pp. 263–284). Oxford University Press. doi:10.1093/acprof:oso/9780199670000.003.0013. (Cit. on p. 54)

Cowen, A. S., Chun, M. M. & Kuhl, B. A. (2014). Neural portraits of perception: Reconstructing face images from evoked brain activity. *NeuroImage*, *94*, 12–22. doi:10.1016/j.neuroimage.2014.03.018. (Cit. on pp. 20, 21, 116, 132, 146)

Crosson, C. W. & Robertson-Tchabo, E. A. (1983). Age and preference for complexity among manifestly creative women. *Human Development*, *26*(3), 149–155. doi:10.1159/000272878. (Cit. on pp. 15, 16, 27, 44, 70)

Cutting, J. E. & Garvin, J. J. (1987). Fractal curves and complexity. *Perception & Psychophysics*, *42*(4), 365–370. doi:10.3758/bf03203093. (Cit. on pp. 45, 75)

Dakin, S. & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, *48*(3), 497–507. doi:10.1016/j.neuron.2005.10.018. (Cit. on pp. 10, 66, 70)

Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. IEEE. doi:10.1109/cvpr.2005.177. (Cit. on p. 81)

de Lurdes Carrito, M., dos Santos, I. M. B., Lefevre, C. E., Whitehead, R. D., da Silva, C. F. & Perrett, D. I. (2016). The role of sexually dimorphic skin colour and shape in attractiveness of male faces. *Evolution and Human Behavior*, *37*(2), 125–133. doi:10.1016/j.evolhumbehav.2015.09.006. (Cit. on p. 147)

de Wit, T. C., Schlooz, W. A., Hulstijn, W. & van Lier, R. (2007). Visual completion and complexity of visual shape in children with pervasive developmental disorder. *European Child & Adolescent Psychiatry*, *16*(3), 168–177. doi:10.1007/s00787-006-0585-9. (Cit. on p. 10)

de Wit, T. C., Bauer, M., Oostenveld, R., Fries, P. & van Lier, R. (2006). Cortical responses to contextual influences in amodal completion. *NeuroImage*, *32*(4), 1815–1825. doi:10.1016/j.neuroimage.2006.05.008. (Cit. on p. 74)

Donderi, D. & McFadden, S. (2005). Compressed file length predicts search time and errors on visual displays. *Displays*, *26*(2), 71–78. doi:10.1016/j.displa.2005.02.002. (Cit. on pp. 4, 7, 9, 34, 45, 74, 75, 147)

Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 73–97. doi:10.1037/0033-2909.132.1.73. (Cit. on pp. 4, 7, 34, 74, 75)

Dong, C., Loy, C. C., He, K. & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307. doi:10.1109/tpami.2015.2439281. (Cit. on p. 113)

Dong, C., Loy, C. C., He, K. & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *Computer vision – ECCV 2014* (pp. 184–199). Springer International Publishing. doi:10.1007/978-3-319-10593-2_13. (Cit. on p. 113)

Dosovitskiy, A. & Brox, T. (2016). Generating images with perceptual similarity metrics based on deep networks. *CoRR*, *abs/1602.02644*. (Cit. on pp. 135, 141).

Du, C., Du, C. & He, H. (2017). Sharing deep generative representation for perceived image reconstruction from human brain activity. *CoRR*, *abs/1704.07575*. (Cit. on p. 132).

Eickenberg, M., Gramfort, A., Varoquaux, G. & Thirion, B. (2017). Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, *152*, 184–194. doi:10.1016/j.neuroimage.2016.10.001. (Cit. on p. 132)

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E.-J. & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*, *369*(6481), 525–525. doi:10.1038/369525a0. (Cit. on p. 80)

Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*(6), 954–978. doi:10.1080/13506280444000607. (Cit. on p. 103)

Epstein, R. & Kanwisher, N. (1998). *Nature*, *392*(6676), 598–601. doi:10.1038/33402. (Cit. on pp. 80, 102)

Eysenck, H. J. (1941). The empirical determination of an aesthetic formula. *Psychological Review*, *48*(1), 83–92. doi:10.1037/h0062483. (Cit. on p. 45)

Farley, F. H. & Weinstock, C. A. (1980). Experimental aesthetics: Children's complexity preference in original art and photoreproductions. *Bulletin of the Psychonomic Society*, *15*(3), 194–196. doi:10.3758/bf03334506. (Cit. on pp. 14, 26)

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N. & Leonard, J. A. (1956). Stimulus correlates of visual pattern recognition: A probability approach. *Journal of Experimental Psychology*, *51*(1), 1–11. doi:10.1037/h0044302. (Cit. on pp. 9, 74)

Fleurian, R. D., Ben-Tal, O., Müllensiefen, D. & Blackwell, T. (2014). Comparing perceptual and computational complexity for short rhythmic patterns. *Procedia - Social and Behavioral Sciences*, *126*, 111–112. doi:10.1016/j.sbspro.2014.02.333. (Cit. on p. 9)

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J. & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, *102*(1), 49–70. doi:10.1348/000712610x498958. (Cit. on pp. 13, 34, 75)

Forsythe, A., Mulhern, G. & Sawey, M. (2008). Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods*, *40*(1), 116–129. doi:10.3758/brm.40.1.116. (Cit. on pp. 6, 7, 34)

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*(4), 1357–1392. doi:10.1152/physrev.00006.2011. (Cit. on p. 103)

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E. & Penny, W. D. (Eds.). (2006). *Statistical parametric mapping: The analysis of functional brain images* (1st ed.). Academic Press. (Cit. on p. 140).

Gao, X., Wang, N., Tao, D. & Li, X. (2012). Face sketch-photo synthesis and retrieval using sparse representation. *IEEE Transactions on Circuits and Systems for Video Technology*, *22*(8), 1213–1226. doi:10.1109/tcsvt.2012.2198090. (Cit. on p. 112)

Gastal, E. S. L. & Oliveira, M. M. (2011). Domain transform for edge-aware image and video processing. *ACM Transactions on Graphics*, *30*(4), 1. doi:10.1145/2010324.1964964. (Cit. on p. 117)

Gatys, L. A., Ecker, A. S. & Bethge, M. (2015). A neural algorithm of artistic style. *CoRR*, *abs/1508.06576*. (Cit. on p. 113).

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*(1), 167–192. doi:10.1146/annurev.psych.58.110405.085632. (Cit. on p. 6)

Glasser, M. F. & Essen, D. C. V. (2011). Mapping human cortical areas in vivo based on myelin content as revealed by t1- and t2-weighted MRI. *Journal of Neuroscience*, *31*(32), 11597–11616. doi:10.1523/jneurosci.2180-11.2011. (Cit. on p. 85)

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... Essen, D. C. V. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. doi:10.1038/nature18933. (Cit. on pp. 84, 85, 103)

Glenberg, A. (2008). *Learning from data : An introduction to statistical reasoning*. New York: Lawrence Erlbaum Associates. (Cit. on p. 33).

Goesaert, E. & Op de Beeck, H. P. (2013). Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. *Journal of Neuroscience*, *33*(19), 8549–8558. doi:10.1523/jneurosci.1829-12.2013. (Cit. on p. 143)

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. D. Fruyt. & F. Ostendorf (Eds.), *Personality psychology in europe* (Vol. 7, pp. 7–28). Tilburg University Press. (Cit. on p. 45).

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial networks. *CoRR*, *abs/1406.2661*. (Cit. on p. 135).

Gough, H. B., Hall, W. B. & Bradley, P. (1996). Forty years of experience with the barron-welsh art scale. In A. Montuori (Ed.), *Unusual associates: A festschrift for frank barron* (252–301). Hampton Press. (Cit. on p. 44).

Graf, L. K. M. & Landwehr, J. R. (2015). A dual-process perspective on fluency-based aesthetics. *Personality and Social Psychology Review*, *19*(4), 395–410. doi:10.1177/1088868315574978. (Cit. on p. 43)

Graham, D. & Field, D. (2007). Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, *21*(1), 149–164. doi:10.1163/156856807782753877. (Cit. on pp. 6, 12)

Grassberger, P. (1986). Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics*, *25*(9), 907–938. doi:10.1007/bf00668821. (Cit. on p. 7)

Greenberg, D. M., Baron-Cohen, S., Stillwell, D. J., Kosinski, M. & Rentfrow, P. J. (2015). Musical preferences are linked to cognitive styles. *PLOS ONE*, *10*(7), e0131151. doi:10.1371/journal.pone.0131151. (Cit. on pp. 16, 54, 56, 57, 154)

Greene, M. R. & Oliva, A. (2010). High-level aftereffects to global scene properties. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1430–1442. doi:10.1037/a0019058. (Cit. on p. 103)

Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*(1), 96–111. (Cit. on p. 17).

Guo, Q., Zhu, C., Xia, Z., Wang, Z. & Liu, Y. (2017). Attribute-controlled face photo synthesis from simple line drawing. *CoRR*, *abs/1702.02805*. (Cit. on p. 163).

Güçlü, U. & van Gerven, M. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005–10014. doi:10.1523/jneurosci.5023-14.2015. (Cit. on pp. 17, 102, 105, 128, 132)

Güçlü, U. & van Gerven, M. (2017a). Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *NeuroImage*, *145*, 329–336. doi:10.1016/j.neuroimage.2015.12.036. (Cit. on pp. 128, 132)

Güçlü, U. & van Gerven, M. (2017b). Modeling the dynamics of human brain activity with recurrent neural networks. *Frontiers in Computational Neuroscience*, *11*. doi:10.3389/fncom.2017.00007. (Cit. on p. 132)

Güçlü, U. & van Gerven, M. (2014). Unsupervised feature learning improves prediction of human brain activity in response to natural images. *PLoS Computational Biology*, *10*(8), e1003724. doi:10.1371/journal.pcbi.1003724. (Cit. on p. 87)

Güçlü, U. & van Gerven, M. (2013). Unsupervised learning of features for Bayesian decoding in functional magnetic resonance imaging. In *Belgian-dutch conference on machine learning*. (Cit. on p. 133).

Güçlü, U., Thielen, J., Hanke, M. & van Gerven, M. (2016). Brains on beats. In D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon & R. Garnett (Eds.), *Advances in neural information processing systems 29: Annual conference on neural information processing systems 2016, december 5-10, 2016, barcelona, spain* (pp. 2101–2109). (Cit. on pp. 17, 102, 132).

Güçlütürk, Y., Güçlü, U., van Lier, R. & van Gerven, M. (2016a). Convolutional sketch inversion. In *Lecture notes in computer science* (pp. 810–824). Springer International Publishing. doi:10.1007/978-3-319-46604-0_56. (Cit. on p. 133)

Güçlütürk, Y., Güçlü, U., Seeliger, K., Bosch, S., van Lier, R. & van Gerven, M. (2017). Deep adversarial neural decoding. *CoRR*, *abs/1705.07109*. (Cit. on p. 74).

Güçlütürk, Y., Jacobs, R. H.A. H. & van Lier, R. (2016b). Liking versus complexity: Decomposing the inverted u-curve. *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2016.00112. (Cit. on pp. 6, 14, 55, 65, 68, 70, 147)

Haesen, B., Boets, B. & Wagemans, J. (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(2), 701–714. doi:10.1016/j.rasd.2010.11.006. (Cit. on pp. 10, 66, 70)

Hahn, A. C. & Perrett, D. I. (2014). Neural and behavioral responses to attractiveness in adult and infant faces. *Neuroscience & Biobehavioral Reviews*, *46*, 591–603. doi:10.1016/j.neubiorev.2014.08.015. (Cit. on p. 147)

Haxby, J. V. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. doi:10.1126/science.1063736. (Cit. on pp. 19, 132)

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. doi:10.1016/j.neuron.2011.08.026. (Cit. on p. 88)

Hayn-Leichsenring, G. U., Lehmann, T. & Redies, C. (2017). Subjective ratings of beauty and aesthetics: Correlations with statistical image properties in western oil paintings. *i-Perception*, *8*(3), 204166951771547. doi:10.1177/2041669517715474. (Cit. on pp. 6, 13, 162)

Hazenberg, S. J. & van Lier, R. (2013). Afterimage watercolors: An exploration of contour-based afterimage filling-in. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00707. (Cit. on p. 164)

Hazenberg, S. J., Jongsma, M. L. A., Koning, A. & van Lier, R. (2014). Differential familiarity effects in amodal completion: Support from behavioral and electrophysiological measurements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 669–684. doi:10.1037/a0034689. (Cit. on p. 74)

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, *abs/1512.03385*. (Cit. on p. 118).

Herzog, T. R. & Shier, R. L. (2000). Complexity, age, and building preference. *Environment and Behavior*, *32*(4), 557–575. doi:10.1177/00139160021972667. (Cit. on p. 74)

Heyduk, R. G. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. *Perception & Psychophysics*, *17*(1), 84–90. doi:10.3758/bf03204003. (Cit. on pp. 8, 13, 14, 54, 74)

Hochberg, J. & McAlister, E. (1953). A quantitative approach, to figural "goodness". *Journal of Experimental Psychology*, *46*(5), 361–364. doi:10.1037/h0055809. (Cit. on p. 4)

Horikawa, T. & Kamitani, Y. (2017a). Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, *8*, 15037. doi:10.1038/ncomms15037. (Cit. on pp. 132, 156)

Horikawa, T. & Kamitani, Y. (2017b). Hierarchical neural representation of dreamed objects revealed by brain decoding with deep neural network features. *Frontiers in Computational Neuroscience*, *11*. doi:10.3389/fncom.2017.00004. (Cit. on p. 132)

Hox, J. J., Moerbeek, M. & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications, second edition (quantitative methodology series)* (2nd ed.). Routledge. (Cit. on p. 64).

Huang, X. & Paradiso, M. A. (2008). V1 response timing and surface filling-in. *Journal of Neurophysiology*, *100*(1), 539–547. doi:10.1152/jn.00997.2007. (Cit. on p. 164)

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. doi:10.1113/jphysiol.1962.sp006837. (Cit. on p. 17)

Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210–1224. doi:10.1016/j.neuron.2012.10.014. (Cit. on p. 102)

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. doi:10.1038/nature17637. (Cit. on p. 102)

Iakovides, S. A., Iliadou, V. T., Bizeli, V. T., Kaprinis, S. G., Fountoulakis, K. N. & Kaprinis, G. S. (2004). *Annals of General Hospital Psychiatry*, *3*(1), 6. doi:10.1186/1475-2832-3-6. (Cit. on p. 10)

Iizuka, S., Simo-Serra, E. & Ishikawa, H. (2016). Let there be color! *ACM Transactions on Graphics*, *35*(4), 1–11. doi:10.1145/2897824.2925974. (Cit. on p. 113)

Imamoglu, Ç. (2000). Complexity, liking and familiarity: Architecture and non-architecture turkish students' assessments of traditional and modern house facades. *Journal of Environmental Psychology*, *20*(1), 5–16. doi:10.1006/jevp.1999.0155. (Cit. on pp. 14, 26)

Ioffe, S. & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, *abs/1502.03167*. (Cit. on pp. 118, 135).

Itti, L. & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. doi:10.1038/35058500. (Cit. on p. 6)

Jacobsen, T. (2004). Individual and group modelling of aesthetic judgment strategies. *British Journal of Psychology*, *95*(1), 41–56. doi:10.1348/000712604322779451. (Cit. on p. 28)

Jacobsen, T. & Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: Analyses of individual judgments. *Perceptual and Motor Skills*, *95*(3), 755–766. doi:10.2466/pms.2002.95.3.755. (Cit. on pp. 15, 27, 28, 46)

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. doi:10.1016/j.neuroimage.2011.09.015. (Cit. on p. 140)

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R. B., ... Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. *CoRR*, *abs/1408.5093*. (Cit. on p. 140).

Johnson, J., Alahi, A. & Li, F. (2016). Perceptual losses for real-time style transfer and super-resolution. *CoRR*, *abs/1603.08155*. (Cit. on pp. 113, 117, 119, 137).

Kamitani, Y. & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. doi:10.1038/nn1444. (Cit. on p. 132)

Kanwisher, N., McDermott, J. & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311. (Cit. on p. 80).

Kay, K. N., Naselaris, T., Prenger, R. J. & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355. doi:10.1038/nature06713. (Cit. on pp. 104, 132)

Kayaert, G. & Wagemans, J. (2009). Delayed shape matching benefits from simplicity and symmetry. *Vision Research*, *49*(7), 708–717. doi:10.1016/j.visres.2009.01.002. (Cit. on pp. 9, 74)

Khaligh-Razavi, S.-M. & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), e1003915. doi:10.1371/journal.pcbi.1003915. (Cit. on p. 132)

Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, *abs/1412.6980*. (Cit. on pp. 118, 141).

Klapuri, A. & Leider, C. (Eds.). (2011). Proceedings of the 12th international society for music information retrieval conference, ISMIR 2011, miami, florida, usa, october 24-28, 2011, University of Miami.

Koch, M., Denzler, J. & Redies, C. (2010). 1/f2 characteristics and isotropy in the fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs. *PLoS ONE*, *5*(8), e12268. doi:10.1371/journal.pone.0012268. (Cit. on p. 12)

Koffka, K. (1935). *Principles of gestalt psychology* (y First edition). Harcourt, Brace and Company. (Cit. on p. 4).

Koide, N., Kubo, T., Nishida, S., Shibata, T. & Ikeda, K. (2015). Art expertise reduces influence of visual salience on fixation in viewing abstract-paintings. *PLOS ONE*, *10*(2), e0117696. doi:10.1371/journal.pone.0117696. (Cit. on pp. 10, 11)

Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nature Reviews Neuroscience*, *7*(3), 220–231. doi:10.1038/nrn1869. (Cit. on p. 164)

Koning, A. & Lier, R. V. (2005). From interpretation to segmentation. *Psychonomic Bulletin & Review*, *12*(5), 917–924. doi:10.3758/bf03196786. (Cit. on pp. 9, 74)

Kourtzi, Z. & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, *20*(9), 3310–3318. (Cit. on p. 80).

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, *1*(1), 417–446. doi:10.1146/annurev-vision-082114-035447. (Cit. on p. 132)

Kurth, F., Eickhoff, S. B., Schleicher, A., Hoemke, L., Zilles, K. & Amunts, K. (2009). Cytoarchitecture and probabilistic maps of the human posterior insular cortex. *Cerebral Cortex*, *20*(6), 1448–1461. doi:10.1093/cercor/bhp208. (Cit. on p. 85)

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L. & Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in Integrative Neuroscience*, *4*. doi:10.3389/fnint.2010.00129. (Cit. on p. 10)

Köhler, W. (1920). *Die physischen gestalten in ruhe und im stationären zustand: Eine naturphilosophische untersuchung (german edition)* (Softcover reprint of the original 1st ed. 1920). Vieweg+Teubner Verlag. (Cit. on p. 4).

Landwehr, J. R., Labroo, A. A. & Herrmann, A. (2011). Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts. *Marketing Science*, *30*(3), 416–429. doi:10.1287/mksc.1110.0633. (Cit. on p. 34)

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T. & van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition & Emotion*, *24*(8), 1377–1388. doi:10.1080/02699930903485076. (Cit. on p. 139)

Law, E., West, K., Mandel, M. I., Bay, M. & Downie, J. S. (2009). Evaluation of algorithms using games: The case of music tagging. In K. Hirata, G. Tzanetakis & K. Yoshii (Eds.), *Proceedings of the 10th international society for music information retrieval conference, ISMIR 2009, kobe international conference center, kobe, japan, october 26-30, 2009* (pp. 387–392). International Society for Music Information Retrieval. (Cit. on p. 83).

Learned-Miller, E., Huang, G. B., RoyChowdhury, A., Li, H. & Hua, G. (2016). Labeled faces in the wild: A survey. In *Advances in face detection and facial image analysis* (pp. 189–248). Springer International Publishing. doi:10.1007/978-3-319-25958-1_8. (Cit. on pp. 114, 115, 128, 158)

Leder, H. & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments : The aesthetic episode - developments and challenges in empirical aesthetics. *British Journal of Psychology*, *105*(4), 443–464. doi:10.1111/bjop.12084. (Cit. on p. 13)

Leder, H., Belke, B., Oeberst, A. & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489–508. doi:10.1348/0007126042369811. (Cit. on p. 13)

Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A. P., Tejani, A., ... Shi, W. (2016). Photo-realistic single image super-resolution using a generative adversarial network. *CoRR*, *abs/1609.04802*. (Cit. on p. 133).

Lee, H. & Kuhl, B. A. (2016). Reconstructing perceived and retrieved faces from activity patterns in lateral parietal cortex. *Journal of Neuroscience*, *36*(22), 6069–6082. doi:10.1523/jneurosci.4286-15.2016. (Cit. on pp. 21, 146)

Leeuwenberg, E. L. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, *76*(2), 216–220. doi:10.1037/h0027285. (Cit. on pp. 5, 45, 75)

Leeuwenberg, E. L. J. (1971). A perceptual coding language for visual and auditory patterns. *The American Journal of Psychology*, *84*(3), 307. doi:10.2307/1420464. (Cit. on pp. 4, 45)

Li, Y., Savvides, M. & Bhagavatula, V. (2006). Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters. In *2006 IEEE international conference on acoustics speed and signal processing proceedings*. IEEE. doi:10.1109/icassp.2006.1660353. (Cit. on p. 112)

Little, A. C., Jones, B. C., Feinberg, D. R. & Perrett, D. I. (2013). Men's strategic preferences for femininity in female faces. *British Journal of Psychology*, *105*(3), 364–381. doi:10.1111/bjop.12043. (Cit. on p. 147)

Liu, Q., Tang, X., Jin, H., Lu, H. & Ma, S. (2005). A nonlinear approach for face sketch synthesis and recognition. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. IEEE. doi:10.1109/cvpr.2005.39. (Cit. on p. 112)

Liu, W., Tang, X. & Liu, J. (2007). Bayesian tensor inference for sketch-based facial photo hallucination. In M. M. Veloso (Ed.), *IJCAI 2007, proceedings of the 20th international joint conference on artificial intelligence, hyderabad, india, january 6-12, 2007* (pp. 2141–2146). (Cit. on p. 112).

Liu, Z., Luo, P., Wang, X. & Tang, X. (2015). Deep learning face attributes in the wild. In *2015 IEEE international conference on computer vision (ICCV)*. IEEE. doi:10.1109/iccv.2015.425. (Cit. on pp. 114, 128, 140, 158)

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. doi:10.1109/tit.1982.1056489. (Cit. on pp. 32, 60)

Ma, D. S., Correll, J. & Wittenbrink, B. (2015). The chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. doi:10.3758/s13428-014-0532-5. (Cit. on p. 139)

Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J. & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, *160*, 43–57. doi:10.1016/j.actpsy.2015.06.005. (Cit. on p. 75)

Mackinnon, D. W. (1962). The nature and nurture of creative talent. *American Psychologist*, *17*(7), 484–495. doi:10.1037/h0046541. (Cit. on p. 44)

Madison, G. & Schiölde, G. (2017). Repeated listening increases the liking for music regardless of its complexity: Implications for the appreciation and aesthetics of music. *Frontiers in Neuroscience*, *11*. doi:10.3389/fnins.2017.00147. (Cit. on pp. 9, 54)

Madsen, S. T. & Widmer, G. (2006). Music complexity measures predicting the listening experience. In *International conference on music perception and cognition*. (Cit. on p. 9).

Mallon, B., Redies, C. & Hayn-Leichsenring, G. U. (2014). Beauty in abstract paintings: Perceptual contrast and statistical properties. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00161. (Cit. on p. 28)

Mandelbrot, B. B. (1977). *Fractals : Form, chance, and dimension*. W. H. Freeman. (Cit. on pp. 13, 45).

Marin, M. M. (2015). Crossing boundaries: Toward a general model of neuroaesthetics. *Frontiers in Human Neuroscience*, *9*. doi:10.3389/fnhum.2015.00443. (Cit. on pp. 71, 163)

Marin, M. M. & Leder, H. (2016). Effects of presentation duration on measures of complexity in affective environmental scenes and representational paintings. *Acta Psychologica*, *163*, 38–58. doi:10.1016/j.actpsy.2015.10.002. (Cit. on pp. 7, 13, 31, 54, 74)

Marin, M. M. & Leder, H. (2013). Examining complexity across domains: Relating subjective and objective measures of affective environmental scenes, paintings and music. *PLoS ONE*, *8*(8), e72412. doi:10.1371/journal.pone.0072412. (Cit. on pp. 6–9, 13, 14, 16, 34, 45, 54, 70, 74, 75, 85, 104)

Marques, G., Domingues, M. A., Langlois, T. & Gouyon, F. (2011). Three current issues in music autotagging. In A. Klapuri & C. Leider (Eds.), *Proceedings of the 12th international society for music information retrieval conference, ISMIR 2011, miami, florida, usa, october 24-28, 2011* (pp. 795–800). University of Miami. (Cit. on p. 83).

Marsik, L., Pokorný, J. & Ilcík, M. (2014). Towards a harmonic complexity of musical pieces. In K. Richta, V. Snásel & J. Pokorný (Eds.), *Proceedings of the dateso 2014 annual international workshop on databases, texts, specifications and objects, roudnice nad labem, czech republic, april 16, 2014.* (Vol. 1139, pp. 1–12). CEUR Workshop Proceedings. CEUR-WS.org. (Cit. on p. 8).

Martinez, A. & Benavente, R. (1998). *The ar face database*. Autonomous University of Barcelona. (Cit. on p. 116).

Martínez-Soto, J., Gonzales-Santos, L., Pasaye, E. & Barrios, F. A. (2013). Exploration of neural correlates of restorative environment exposure through functional magnetic resonance. *Intelligent Buildings International*, *5*(sup1), 10–28. doi:10.1080/17508975.2013.807765. (Cit. on p. 105)

Mauch, M. & Levy, M. (2011). Structural change on multiple time scales as a correlate of musical complexity. In A. Klapuri & C. Leider (Eds.), *Proceedings of the 12th international society for music information retrieval conference, ISMIR 2011, miami, florida, usa, october 24-28, 2011* (pp. 489–494). University of Miami. (Cit. on p. 8).

Mavrides, C. M. & Brown, D. R. (1969). Discrimination and reproduction of patterns: Feature measures and constraint redundancy as predictors. *Perception & Psychophysics*, *6*(5), 276–280. doi:10.3758/bf03210098. (Cit. on pp. 9, 74)

McManus, I. C., Cook, R. & Hunt, A. (2010). Beyond the golden section and normative aesthetics: Why do individuals differ so much in their aesthetic preferences for rectangles? *Psychology of Aesthetics, Creativity, and the Arts*, *4*(2), 113–126. doi:10.1037/a0017316. (Cit. on pp. 15, 27)

McWhinnie, H. J. (1968). A review of research on aesthetic measure. *Acta Psychologica*, *28*, 363–375. doi:10.1016/0001-6918(68)90025-5. (Cit. on p. 44)

Messer, K., anad J. Kittler, J. M., Luettin, J. & Maître, G. (1999). Xm2vtsdb: The extended m2vts database. In *Audio- and video-based biometric person authentication*. (Cit. on p. 116).

Meyer, L. B. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism*, *15*(4), 412. doi:10.2307/427154. (Cit. on p. 8)

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A. & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191–1195. doi:10.1126/science.1152876. (Cit. on pp. 104, 132)

Mitsa, T. (2010). *Temporal data mining*. Boca Raton, FL: Chapman & Hall/CRC. (Cit. on p. 33).

Miyawaki, Y., Uchida, H., Yamashita, O., aki Sato, M., Morito, Y., Tanabe, H. C., ... Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, *60*(5), 915–929. doi:10.1016/j.neuron.2008.11.004. (Cit. on pp. 19, 132)

Moerel, M., Martino, F. D. & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, *8*. doi:10.3389/fnins.2014.00225. (Cit. on p. 85)

Moffett, A. (1969). Stimulus complexity as a determinant of visual attention in infants. *Journal of Experimental Child Psychology*, *8*(1), 173–179. doi:10.1016/0022-0965(69)90038-1. (Cit. on p. 9)

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T. & Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, *13*(4), 684–701. doi:10.1006/nimg.2000.0715. (Cit. on p. 85)

Morosan, P., Schleicher, A., Amunts, K. & Zilles, K. (2005). Multimodal architectonic mapping of human superior temporal gyrus. *Anatomy and Embryology*, *210*(5-6), 401–406. doi:10.1007/s00429-005-0029-1. (Cit. on p. 85)

Mumford, J. A., Turner, B. O., Ashby, F. G. & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643. doi:10.1016/j.neuroimage.2011.08.076. (Cit. on p. 140)

Munsinger, H., Kessen, W. & Kessen, M. L. (1964). Age and uncertainty: Developmental variation in preference for variability. *Journal of Experimental Child Psychology*, *1*(1), 1–15. doi:10.1016/0022-0965(64)90002-5. (Cit. on pp. 16, 44, 70)

Muth, C. & Carbon, C.-C. (2013). The aesthetic aha: On the pleasure of having insights into gestalt. *Acta Psychologica*, *144*(1), 25–30. doi:10.1016/j.actpsy.2013.05.001. (Cit. on p. 16)

Muth, C., Raab, M. H. & Carbon, C.-C. (2016). Semantic stability is more pleasurable in unstable episodic contexts. on the relevance of perceptual challenge in art appreciation. *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2016.00043. (Cit. on p. 13)

Muth, C., Raab, M. H. & Carbon, C.-C. (2015). The stream of experience when watching artistic movies. dynamic aesthetic effects revealed by the continuous evaluation procedure (CEP). *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00365. (Cit. on pp. 13, 74, 86)

Nadal, M., Munar, E., Marty, G. & Cela-Conde, C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, *28*(2), 173–191. doi:10.2190/em.28.2.d. (Cit. on pp. 14, 15, 26, 74, 75, 152)

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M. & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*(6), 902–915. doi:10.1016/j.neuron.2009.09.006. (Cit. on pp. 20, 132, 133)

Naselaris, T., Kay, K. N., Nishimoto, S. & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410. doi:10.1016/j.neuroimage.2010.07.073. (Cit. on pp. 76, 104, 132)

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B. & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, *21*(19), 1641–1646. doi:10.1016/j.cub.2011.08.031. (Cit. on pp. 20, 132)

North, A. C. (2010). Individual differences in musical taste. *The American Journal of Psychology*, *123*(2), 199. doi:10.5406/amerjpsyc.123.2.0199. (Cit. on p. 68)

North, A. C. & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: A Journal of Research in Music Cognition*, *14*(1-2), 77–93. doi:10.1037/h0094090. (Cit. on pp. 14, 54, 74)

Nusbaum, E. C. & Silvia, P. J. (2010). Shivers and timbres. *Social Psychological and Personality Science*, *2*(2), 199–204. doi:10.1177/1948550610386810. (Cit. on pp. 14, 27)

Oliva, A. & Torralba, A. (2001). *International Journal of Computer Vision*, *42*(3), 145–175. doi:10.1023/a:1011139631724. (Cit. on pp. 6, 103)

Orr, M. G. (2005). Relationship between complexity and liking as a function of expertise. *Music Perception: An Interdisciplinary Journal*, *22*(4), 583–611. doi:10.1525/mp.2005.22.4.583. (Cit. on pp. 8, 13, 54, 74)

Orr, M. G. & Ohlsson, S. (2001). The relationship between musical complexity and liking in jazz and bluegrass. *Psychology of Music*, *29*(2), 108–127. doi:10.1177/0305735601292002. (Cit. on pp. 8, 54)

Palmer, S. E. & Griscom, W. S. (2012). Accounting for taste: Individual differences in preference for harmony. *Psychonomic Bulletin & Review*, *20*(3), 453–461. doi:10.3758/s13423-012-0355-2. (Cit. on pp. 14, 27, 46)

Pandya, D. N. & Sanides, F. (1973). Architectonic parcellation of the temporal operculum in rhesus monkey and its projection pattern. *Zeitschrift fr Anatomie und Entwicklungsgeschichte*, *139*(2), 127–161. doi:10.1007/bf00523634. (Cit. on p. 85)

Park, S., Brady, T. F., Greene, M. R. & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience*, *31*(4), 1333–1340. doi:10.1523/jneurosci.3885-10.2011. (Cit. on p. 102)

Parkhi, O. M., Vedaldi, A. & Zisserman, A. (2015). Deep face recognition. In *Procedings of the british machine vision conference 2015*. British Machine Vision Association. doi:10.5244/c.29.41. (Cit. on p. 135)

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T. & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. *CoRR*, *abs/1604.07379*. (Cit. on p. 133).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830. (Cit. on p. 140).

Percino, G., Klimek, P. & Thurner, S. (2014). Instrumentational complexity of music genres and why simplicity sells. *PLoS ONE*, *9*(12), e115255. doi:10.1371/journal.pone.0115255. (Cit. on pp. 8, 12, 17, 57)

Perrett, D. I., May, K. A & Yoshikawa, S. (1994). Facial shape and judgements of female attractiveness. *Nature*, *368*(6468), 239–242. doi:10.1038/368239a0. (Cit. on p. 147)

Petersen, K. B. & Pedersen, M. S. (2012). *The matrix cookbook*. Technical University of Denmark. (Cit. on p. 138).

Phelps, M., Kuhl, D. & Mazziota, J. (1981). Metabolic mapping of the brain's response to visual stimulation: Studies in humans. *Science*, *211*(4489), 1445–1448. doi:10.1126/science.6970412. (Cit. on pp. 17, 76)

Pinna, B., Brelstaff, G. & Spillmann, L. (2001). Surface color from boundaries: A new 'watercolor' illusion. *Vision Research*, *41*(20), 2669–2676. doi:10.1016/s0042-6989(01)00105-5. (Cit. on p. 164)

Potter, R. F. & Choi, J. (2006). The effects of auditory structural complexity on attitudes, attention, arousal, and memory. *Media Psychology*, *8*(4), 395–419. doi:10.1207/s1532785xmep0804_4. (Cit. on pp. 9, 74)

Pugach, C., Leder, H. & Graham, D. J. (2017). How stable are human aesthetic preferences across the lifespan? *Frontiers in Human Neuroscience*, *11*. doi:10.3389/fnhum.2017.00289. (Cit. on p. 70)

Radford, A., Metz, L. & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, *abs/1511.06434*. (Cit. on pp. 135, 141, 144).

Ramanujan, S. (1914). Modular equations and approximations to π. *Quarterly Journal of Mathematics*, *45*, 350–372. (Cit. on pp. vii, 3).

Reber, R., Schwarz, N. & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364–382. doi:10.1207/s15327957pspr0804_3. (Cit. on pp. 43, 66)

Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Frontiers in Human Neuroscience*, *09*. doi:10.3389/fnhum.2015.00218. (Cit. on p. 13)

Redies, C., Amirshahi, S. A., Koch, M. & Denzler, J. (2012). PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects. In *Computer vision – ECCV 2012. workshops and demonstrations* (pp. 522–531). Springer Berlin Heidelberg. doi:10.1007/978-3-642-33863-2_54. (Cit. on pp. 13, 75, 80, 81)

Renninger, L. W. & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, *44*(19), 2301–2311. doi:10.1016/j.visres.2004.04.006. (Cit. on p. 103)

Rentfrow, P. J., Goldberg, L. R., Stillwell, D. J., Kosinski, M., Gosling, S. D. & Levitin, D. J. (2012). The song remains the same: A replication and extension of the MUSIC model. *Music Perception: An Interdisciplinary Journal*, *30*(2), 161–185. doi:10.1525/mp.2012.30.2.161. (Cit. on pp. 16, 54, 57)

Rentfrow, P. J., Goldberg, L. R. & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology*, *100*(6), 1139–1157. doi:10.1037/a0022406. (Cit. on pp. 12, 16, 54, 57, 60, 69)

Richards, J. E. (2010). The development of attention to simple and complex visual stimuli in infants: Behavioral and psychophysiological measures. *Developmental Review*, *30*(2), 203–219. doi:10.1016/j.dr.2010.03.005. (Cit. on p. 9)

Robertson, A. E. & Simmons, D. R. (2012). The relationship between sensory sensitivity and autistic traits in the general population. *Journal of Autism and Developmental Disorders*, *43*(4), 775–784. doi:10.1007/s10803-012-1608-7. (Cit. on p. 70)

Rokach, L. (2009). A survey of clustering algorithms. In *Data mining and knowledge discovery handbook* (pp. 269–298). Springer US. doi:10.1007/978-0-387-09823-4_14. (Cit. on p. 32)

Roncaglia-Denissen, M. P., Roor, D. A., Chen, A. & Sadakata, M. (2016). The enhanced musical rhythmic perception in second language learners. *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2016.00288. (Cit. on p. 10)

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. doi:10.1016/0377-0427(87)90125-7. (Cit. on pp. 32, 37, 61)

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252. doi:10.1007/s11263-015-0816-y. (Cit. on p. 79)

Saklofske, D. H. (1975). Aesthetic complexity and exploratory behavior. *Perceptual and Motor Skills*, *41*(2), 363–368. doi:10.2466/pms.1975.41.2.363. (Cit. on pp. 14, 26)

Sallavanti, M. I., Szilagyi, V. E. & Crawley, E. J. (2015). The role of complexity in music uses. *Psychology of Music*, *44*(4), 757–768. doi:10.1177/0305735615591843. (Cit. on pp. 8, 9)

Sambrook, T. & Whiten, A. (1997). On the nature of complexity in cognitive and behavioural science. *Theory & Psychology*, *7*(2), 191–213. doi:10.1177/0959354397072004. (Cit. on pp. 4, 7)

Samson, F., Hyde, K. L., Bertone, A., Soulières, I., Mendrek, A., Ahad, P., ... Zeffiro, T. A. (2011a). Atypical processing of auditory temporal complexity in autistics. *Neuropsychologia*, *49*(3), 546–555. doi:10.1016/j.neuropsychologia.2010.12.033. (Cit. on p. 18)

Samson, F., Zeffiro, T. A., Toussaint, A. & Belin, P. (2011b). Stimulus complexity and categorical effects in human auditory cortex: An activation likelihood estimation meta-analysis. *Frontiers in Psychology*, *1*. doi:10.3389/fpsyg.2010.00241. (Cit. on pp. 9, 18)

Sangkloy, P., Lu, J., Fang, C., Yu, F. & Hays, J. (2016). Scribbler: Controlling deep image synthesis with sketch and color. *CoRR*, *abs/1612.00835*. (Cit. on p. 163).

Schiffman, H. R. & Bobko, D. J. (1974). Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology*, *103*(1), 156–159. doi:10.1037/h0036794. (Cit. on pp. 9, 74)

Schlaug, G. (2006). The brain of musicians. *Annals of the New York Academy of Sciences*, *930*(1), 281–299. doi:10.1111/j.1749-6632.2001.tb05739.x. (Cit. on p. 10)

Schlochtermeier, L. H., Kuchinke, L., Pehrs, C., Urton, K., Kappelhoff, H. & Jacobs, A. M. (2013). Emotional picture and word processing: An fMRI study on effects of stimulus complexity. *PLoS ONE*, *8*(2), e55619. doi:10.1371/journal.pone.0055619. (Cit. on p. 18)

Schoenmakers, S., Güçlü, U., van Gerven, M. & Heskes, T. (2015). Gaussian mixture models and semantic gating improve reconstructions from human brain activity. *Frontiers in Computational Neuroscience*, *8*. doi:10.3389/fncom.2014.00173. (Cit. on p. 133)

Schoenmakers, S., Barth, M., Heskes, T. & van Gerven, M. (2013). Linear reconstruction of perceived images from human brain activity. *NeuroImage*, *83*, 951–961. doi:10.1016/j.neuroimage.2013.07.043. (Cit. on pp. 19, 133, 146)

Sereno, M., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T., ... Tootell, R. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893. doi:10.1126/science.7754376. (Cit. on p. 80)

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x. (Cit. on pp. 2, 74)

Shier, J. (2011). Filling space with random fractal non-overlapping simple shapes. In *International society of the arts, mathematics, and architecture*. (Cit. on pp. 17, 29, 152).

Shier, J. & Bourke, P. (2013). An algorithm for random fractal filling of space. *Computer Graphics Forum*, *32*(8), 89–97. doi:10.1111/cgf.12163. (Cit. on p. 29)

Shigeto, H., Ishiguro, J. & Nittono, H. (2011). Effects of visual stimulus complexity on event-related brain potentials and viewing duration in a free-viewing task. *Neuroscience Letters*, *497*(2), 85–89. doi:10.1016/j.neulet.2011.04.035. (Cit. on pp. 18, 76)

Shmulevich, I. & Povel, D.-J. (2000). Complexity measures of musical rhythms. In P. Desain & L. Windsor (Eds.), *Rhythm perception and production* (pp. 239–244). Lisse, NL: Swets & Zeitlinger. (Cit. on pp. 2, 8, 74, 75).

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. doi:10.1037/0033-2909.86.2.420. (Cit. on p. 61)

Silva, M. P. D., Courboulay, V. & Estraillier, P. (2011). Image complexity measure based on visual attention. In *2011 18th IEEE international conference on image processing*. IEEE. doi:10.1109/icip.2011.6116371. (Cit. on p. 6)

Silvia, P. J. (2005). Cognitive appraisals and interest in visual art: Exploring an appraisal theory of aesthetic emotions. *Empirical Studies of the Arts*, *23*(2), 119–133. doi:10.2190/12av-ah2p-mceh-289e. (Cit. on p. 26)

Simo-Serra, E., Iizuka, S., Sasaki, K. & Ishikawa, H. (2016). Learning to simplify. *ACM Transactions on Graphics*, *35*(4), 1–11. doi:10.1145/2897824.2925972. (Cit. on p. 113)

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, *79*(5), 369–382. doi:10.1037/h0033118. (Cit. on pp. 9, 74)

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, *abs/1409.1556*. (Cit. on pp. 119, 137).

Sitzer, M., Diehl, R. R. & Hennerici, M. (1992). Visually evoked cerebral blood flow responses. *Journal of Neuroimaging*, *2*(2), 65–70. doi:10.1111/jon19922265. (Cit. on pp. 17, 18, 76)

Solomonoff, R. (1986). The application of algorithmic probability to problems in artificial intelligence. In *Uncertainty in artificial intelligence* (pp. 473–491). Elsevier. doi:10.1016/b978-0-444-70058-2.50040-1. (Cit. on pp. 4, 34, 81)

Spehar, B., Wong, S., van de Klundert, S., Lui, J., Clifford, C. W. G. & Taylor, R. P. (2015). Beauty and the beholder: The role of visual sensitivity in visual preference. *Frontiers in Human Neuroscience*, *9*. doi:10.3389/fnhum.2015.00514. (Cit. on pp. 6, 15, 16, 46, 71, 162)

Spehar, B., Walker, N. & Taylor, R. P. (2016). Taxonomy of individual variations in aesthetic responses to fractal patterns. *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2016.00350. (Cit. on pp. 55, 162)

Spehar, B., Clifford, C. W., Newell, B. R. & Taylor, R. P. (2003). Universal aesthetic of fractals. *Computers & Graphics*, *27*(5), 813–820. doi:10.1016/s0097-8493(03)00154-7. (Cit. on pp. 13, 45, 75)

Stamps, A. E. (2004). Mystery, complexity, legibility and coherence: A meta-analysis. *Journal of Environmental Psychology*, *24*(1), 1–16. doi:10.1016/s0272-4944(03)00023-9. (Cit. on p. 74)

Streich, S. (2007). *Music complexity: A multi-faceted description of audio content* (Doctoral dissertation, Pompeu Fabra University). (Cit. on pp. 8, 75).

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C. & Heagins, T. B. (2015). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods*, *48*(3), 1197–1204. doi:10.3758/s13428-015-0641-9. (Cit. on p. 139)

Strom, M. A., Zebrowitz, L. A., Zhang, S., Bronstad, P. M. & Lee, H. K. (2012). Skin and bones: The contribution of skin tone and facial structure to racial prototypicality ratings. *PLoS ONE*, *7*(7), e41193. doi:10.1371/journal.pone.0041193. (Cit. on p. 147)

Tagg, P. (1982). Analysing popular music: Theory, method and practice. *Popular Music*, *2*, 37. doi:10.1017/s0261143000001227. (Cit. on p. 12)

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*(1), 109–139. doi:10.1146/annurev.ne.19.030196.000545. (Cit. on p. 17)

Tang, X. & Wang, X. (2003). Face sketch synthesis and recognition. In *Proceedings ninth IEEE international conference on computer vision*. IEEE. doi:10.1109/iccv.2003.1238414. (Cit. on p. 112)

Taylor, R., Newell, B., Spehar, B. & Clifford, C. (2005). Fractals: A resonance between art and nature. In *Mathematics and culture II* (pp. 53–63). Springer-Verlag. doi:10.1007/3-540-26443-4_6. (Cit. on pp. 6, 75)

Taylor, R. P., Micolich, A. P. & Jonas, D. (1999). *Nature*, *399*(6735), 422–422. doi:10.1038/20833. (Cit. on p. 13)

Thaler, L., Schütz, A., Goodale, M. & Gegenfurtner, K. (2013). What is the best fixation target? the effect of target shape on stability of fixational eye movements. *Vision Research*, *76*, 31–42. doi:10.1016/j.visres.2012.10.012. (Cit. on p. 139)

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D. & Dehaene, S. (2006). Inverse retinotopy: Inferring the visual content of images from brain activation patterns. *NeuroImage*, *33*(4), 1104–1116. doi:10.1016/j.neuroimage.2006.06.062. (Cit. on pp. 19, 132, 133)

Thorndike, E. L. (1917). Individual differences in judgments of the beauty of simple forms. *Psychological Review*, *24*(2), 147–153. doi:10.1037/h0073175. (Cit. on pp. 14, 15)

Thul, E. & Toussaint, G. T. (2008). Rhythm complexity measures: A comparison of mathematical models of human perception and performance. In J. P. Bello, E. Chew & D. Turnbull (Eds.), *ISMIR 2008, 9th international conference on music information retrieval, drexel university, philadelphia, pa, usa, september 14-18, 2008* (pp. 663–668). (Cit. on p. 8).

Tokui, S., Oono, K., Hido, S. & Clayton, J. (2015). Chainer: A next-generation open source framework for deep learning. In *Machine learning systems*. (Cit. on pp. 118, 140).

Triarhou, L. C. (2007a). A proposed number system for the 107 cortical areas of economo and koskinas, and brodmann area correlations. *Stereotactic and Functional Neurosurgery*, *85*(5), 204–215. doi:10.1159/000103259. (Cit. on p. 85)

Triarhou, L. C. (2007b). The economo-koskinas atlas revisited: Cytoarchitectonics and functional context. *Stereotactic and Functional Neurosurgery*, *85*(5), 195–203. doi:10.1159/000103258. (Cit. on p. 85)

Tuch, A. N., Bargas-Avila, J. A., Opwis, K. & Wilhelm, F. H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies*, *67*(9), 703–715. doi:10.1016/j.ijhcs.2009.04.002. (Cit. on p. 7)

Tveit, M., Ode, Å. & Fry, G. (2006). Key concepts in a framework for analysing visual landscape character. *Landscape Research*, *31*(3), 229–255. doi:10.1080/01426390600783269. (Cit. on p. 74)

van Gerven, M. (2017). A primer on encoding models in sensory neuroscience. *Journal of Mathematical Psychology*, *76*, 172–183. doi:10.1016/j.jmp.2016.06.009. (Cit. on p. 132)

van Gerven, M. & Heskes, T. (2012). A linear gaussian framework for decoding of perceived images. In *2012 second international workshop on pattern recognition in NeuroImaging*. IEEE. doi:10.1109/prni.2012.10. (Cit. on p. 146)

van Gerven, M., de Lange, F. P. & Heskes, T. (2010). Neural decoding with hierarchical generative models. *Neural Computation*, *22*(12), 3127–3142. doi:10.1162/neco_a_00047. (Cit. on pp. 19, 132)

Van de Cruys, S. & Wagemans, J. (2011). Putting reward in art: A tentative prediction error account of visual art. *i-Perception*, *2*(9), 1035–1062. doi:10.1068/i0466aap. (Cit. on p. 16)

van der Helm, P. A. (2014). *Simplicity in vision: A multidisciplinary account of perceptual organization*. Cambridge University Press. (Cit. on pp. 5, 75).

van der Helm, P. A., van Lier, R. J. & Leeuwenberg, E. L. J. (1992). Serial pattern complexity: Irregularity and hierarchy. *Perception*, *21*(4), 517–544. (Cit. on p. 45).

van der Helm, P. & Leeuwenberg, E. (1991). Accessibility: A criterion for regularity and hierarchy in visual pattern codes. *Journal of Mathematical Psychology*, *35*(2), 151–213. doi:10.1016/0022-2496(91)90025-o. (Cit. on p. 5)

van Lier, R. (1998). Effects of physical connectivity on the representational unity of multi-part configurations. *Cognition*, *69*(1), B1–B9. doi:10.1016/s0010-0277(98)00055-9. (Cit. on p. 9)

van Lier, R. (1999). Investigating global effects in visual occlusion: From a partly occluded square to the back of a tree-trunk. *Acta Psychologica*, *102*(2-3), 203–220. doi:10.1016/s0001-6918(98)00055-9. (Cit. on p. 74)

van Lier, R. (1996). *Simplicity of visual shape: A structural information approach* (Doctoral dissertation, University of Nijmegen). (Cit. on p. 5).

van Lier, R. (2001). Simplicity, regularity, and perceptual interpretations: A structural information approach. In *From fragments to objects - segmentation and grouping in vision* (pp. 331–352). Elsevier. doi:10.1016/s0166-4115(01)80031-5. (Cit. on pp. 5, 75)

van Lier, R., Vergeer, M. & Anstis, S. (2009). Filling-in afterimage colors between the lines. *Current Biology*, *19*(8), R323–R324. doi:10.1016/j.cub.2009.03.010. (Cit. on p. 164)

van Lier, R., van der Helm, P. & Leeuwenberg, E. (1994). Integrating global and local aspects of visual occlusion. *Perception*, *23*(8), 883–903. doi:10.1068/p230883. (Cit. on p. 5)

Vanmarcke, S., Noens, I., Steyaert, J. & Wagemans, J. (2017). Spatial frequency priming of scene perception in adolescents with and without ASD. *Journal of Autism and Developmental Disorders*, *47*(7), 2023–2038. doi:10.1007/s10803-017-3123-3. (Cit. on p. 10)

Vergeer, M., Anstis, S. & van Lier, R. (2015). Flexible color perception depending on the shape and positioning of achromatic contours. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00620. (Cit. on p. 128)

Vitz, P. C. (1966). Preference for different amounts of visual complexity. *Behavioral Science*, *11*(2), 105–114. doi:10.1002/bs.3830110204. (Cit. on pp. 14, 15, 26, 27)

Vogt, S. & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception*, *36*(1), 91–100. doi:10.1068/p5262. (Cit. on pp. 10, 11)

Vrins, S., Hunnius, S. & van Lier, R. (2011). Volume completion in 4.5-month-old infants. *Acta Psychologica*, *138*(1), 92–99. doi:10.1016/j.actpsy.2011.05.010. (Cit. on p. 9)

Vuust, P. & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.01111. (Cit. on p. 8)

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M. & von der Heydt, R. (2012a). A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. doi:10.1037/a0029333. (Cit. on p. 5)

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A. & van Leeuwen, C. (2012b). A century of gestalt psychology in visual perception: II. conceptual and theoretical foundations. *Psychological Bulletin*, *138*(6), 1218–1252. doi:10.1037/a0029334. (Cit. on p. 5)

Wang, N., Tao, D., Gao, X., Li, X. & Li, J. (2013a). A comprehensive survey to face hallucination. *International Journal of Computer Vision*, *106*(1), 9–30. doi:10.1007/s11263-013-0645-9. (Cit. on p. 127)

Wang, N., Tao, D., Gao, X., Li, X. & Li, J. (2013b). Transductive face sketch-photo synthesis. *IEEE Transactions on Neural Networks and Learning Systems*, *24*(9), 1364–1376. doi:10.1109/tnnls.2013.2258174. (Cit. on p. 112)

Wang, X. & Tang, X. (2009). Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(11), 1955–1967. doi:10.1109/tpami.2008.222. (Cit. on pp. 112, 114–116, 123, 127)

Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. doi:10.1109/tip.2003.819861. (Cit. on pp. 121, 141)

Weiss, C. & Muller, M. (2015). Tonal complexity features for style classification of classical music. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. doi:10.1109/icassp.2015.7178057. (Cit. on p. 8)

Wertheimer, M. (1912). Experimentelle studien über das sehen von bewegung. *Zeitschrift für Psychologie*, *61*(1), 161–265. (Cit. on p. 4).

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., ... Wakabayashi, A. (2006). Predicting autism spectrum quotient (AQ) from the systemizing quotient-revised (SQ-r) and empathy quotient (EQ). *Brain Research*, *1079*(1), 47–56. doi:10.1016/j.brainres.2006.01.012. (Cit. on pp. 58, 59)

Wilkins, R. W., Hodges, D. A., Laurienti, P. J., Steen, M. & Burdette, J. H. (2014). Network science and the effects of music preference on functional brain connectivity: From beethoven to eminem. *Scientific Reports*, *4*(1). doi:10.1038/srep06130. (Cit. on pp. 19, 76)

Xiao, B., Gao, X., Tao, D. & Li, X. (2009). A new approach for face recognition by sketches in photos. *Signal Processing*, *89*(8), 1576–1588. doi:10.1016/j.sigpro.2009.02.008. (Cit. on p. 112)

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D. & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. doi:10.1073/pnas.1403112111. (Cit. on p. 132)

Yamins, D. L. K. & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365. doi:10.1038/nn.4244. (Cit. on p. 132)

Yovel, G. & Freiwald, W. A. (2013). Face recognition systems in monkey and human: Are they the same thing? *F1000Prime Reports*, *5*. doi:10.12703/p5-10. (Cit. on p. 20)

Zeki, S., Romaya, J. P., Benincasa, D. M. T. & Atiyah, M. F. (2014). The experience of mathematical beauty and its neural correlates. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00068. (Cit. on p. 105)

Zhang, L., Lin, L., Wu, X., Ding, S. & Zhang, L. (2015). End-to-end photo-sketch generation via fully convolutional representation learning. In *Proceedings of the 5th ACM on international conference on multimedia retrieval - ICMR '15*. ACM Press. doi:10.1145/2671188.2749321. (Cit. on pp. 112, 113)

Zhang, S., Gao, X., Wang, N. & Li, J. (2016). Robust face sketch style synthesis. *IEEE Transactions on Image Processing*, *25*(1), 220–232. doi:10.1109/tip.2015.2501755. (Cit. on pp. 112, 113, 126, 133)

Zhang, W., Wang, X. & Tang, X. (2010). Lighting and pose robust face sketch synthesis. In *Computer vision – ECCV 2010* (pp. 420–433). Springer Berlin Heidelberg. doi:10.1007/978-3-642-15567-3_31. (Cit. on p. 112)

Zubin, J. (1938). A technique for measuring like-mindedness. *The Journal of Abnormal and Social Psychology*, *33*(4), 508–516. doi:10.1037/h0055441. (Cit. on p. 27)

Nederlandse samenvatting

In dit proefschrift, staan de effecten van complexiteit op verschillende aspecten van de menselijke waarneming centraal. De waardering van stimulus en de rol van complexiteit in de kunst is daarbij van bijzonder belang is geweest (bijv. In Hoofdstukken 2, 3, 4 en 5). Daarnaast hebben we ook de rol van complexiteit bij reconstructies van de waarneming onderzocht. (bijv. in hoofdstuk 4 en 6).

In Hoofdstuk 2 hebben we individuele verschillen in complexiteitsvoorkeuren van mensen onderzocht op basis van hun waardering voor abstracte digitale kunst. De belangrijkste resultaten van dit hoofdstuk zijn als volgt: (i) We hebben twee groepen deelnemers geïndentificeerd die duidelijk tegenovergestelde complexiteitsvoorkeuren hadden voor een gecontroleerde set achromatische geometrische patronen. (ii) We hebben aangetoond dat een van de bekendste resultaten met betrekking tot de esthetische voorkeur, de zogenaamde omgekeerde U-curve, een mogelijk artefact is die ontstaat door het selecteren van een niet-ideale analysemethode.

In Hoofdstuk 3, hebben we de ideeën die werden ontwikkeld in hoofdstuk 2 uitgebreid met het onderzoeken naar individuele verschillen met betrekking tot complexiteitsvoorkeuren voor muziek. We hebben aangetoond dat de resultaten met betrekking tot de dualiteit van de complexiteitsvoorkeuren die we in hoofdstuk 2 hebben verkregen, niet beperkt zijn tot het visuele domein en de gebruikte (sterk gecontroleerde stimuli), maar ook duidelijk te zien zijn in het auditieve domein met complexe muziekstimuli.

Hoofdstuk 4 is het laatste hoofdstuk met de focus op de rol van complexiteit tijdens waarneming, en de constructie van het waargenomene. In Hoofdstuk 4 hebben we de neurale representaties van complexiteit in de visuele en auditieve modaliteiten onderzocht. De belangrijkste resultaten van dit hoofdstuk zijn: (i) We hebben neurale kenmerken van complexiteit in de auditieve en visuele sensorische cortex in kaart gebracht. (ii) We hebben de complexiteitsgevoeligheid van individuele interessegebieden (ROI's) gekwantificeerd. Dit laatste liet een verandering van gevoeligheid zien (van fijnkorrelig tot grover) in een gradiënt van lagere naar hogere gebieden. (iii) We hebben aangetoond dat de representatie van complexiteit gekenmerkt wordt door een distributie in de parahippocampale gebieden die vergelijkbaar is met, of zelfs beter is dan, de ROI's in de vroege visuele cortex. Deze bevinding ondersteunt het idee dat globale scène-eigenschappen (zoals complexiteit) een belangrijke rol spelen bij scèneverwerking. (iv) We hebben aangetoond dat gebieden van de auditieve cortex die syntactische taalcomplexiteit representeren, ook de muziekcomplexiteit representeren. (v) We hebben aangetoond dat vroege sensorische cortices reageren op alle complexiteitsmetingen, terwijl voxels in de ROI's van associatieve sensorische cortices meer gespecialiseerd worden langs de sensorische hiërarchie.

In hoofdstuk 5 stelden we ons de vraag of het mogelijk zou zijn om een complexe stimulus, zoals een foto van een gezicht, te reconstrueren uit de eenvoudigere weergave ervan, d.w.z. een schets. We hebben aangetoond dat dit inderdaad mogelijk is met behulp van geavanceerde deep-learning technieken. De belangrijkste resultaten van hoofdstuk 5 zijn als

volgt: (i) We hebben deep-learning neurale netwerkmodellen ontwikkeld die gezichten kunnen synthetiseren uitgaande van ruwe (zwart-wit) schetsen. De resultaten zijn state-of-the art te noemen. (ii) We hebben toepassingen voorgesteld van ons model in de kunsten, kunstgeschiedenis en forensische geneeskunde. (iii) We hebben fotorealistische portretten van beroemde Nederlandse kunstenaars gereconstrueerd. Dit laatste is o.a. gedaan om de mogelijkheden voor kunsthistorische toepassingen en beeldende kunst te demonstreren.

Tot slot hebben we in hoofdstuk 6 een methode ontwikkeld voor het decoderen van gezichten op basis van stimulus-geïnitieerde hersenactiviteit, zoals gemeten met behulp van fMRI. We stelden ons daarbij de vraag of de verschillen in complexiteitsniveaus van gezichtsstimuli de decoderingsprestaties beïnvloeden. De belangrijkste resultaten van hoofdstuk 6 zijn als volgt: (i) We hebben een nieuwe benadering voorgesteld voor het reconstrueren van complexe stimuluspercepten door probabilistische inferenties te combineren met deep learning. (ii) We genereerden state-of-the-art reconstructies van waargenomen gezichten op basis van gemeten hersenactiviteit. (iii) We hebben aangetoond dat stimuluscomplexiteit een belangrijke factor is die de decoderingsprestaties beïnvloedt.

Het is meer dan een eeuw geleden sinds de vroegste studies over de relatie tussen complexiteit en perceptie. Gedurende deze periode zijn er talloze ideeën geweest over de effecten van complexiteit op een grote variëteit aan perceptuele verschijnselen. Er zijn diverse pogingen gedaan om de invloed van complexiteit te formaliseren. Tegenwoordig zijn er diverse nieuwe technieken waardoor een stap voorwaarts gezet kan worden. Vooral het aanpakken van dit probleem vanuit verschillende invalshoeken, zoals computationeel, gedragsmatig en neuraal, zoals we in dit proefschrift hebben omarmd, lijkt een veelbelovende benadering.

Acknowledgements

First and foremost, I would like to thank my supervisors Rob, Richard and Harold, and thesis committee Richard van Wezel, Johan Wagemans and Claus-Christian Carbon.

I have been extremely lucky to have Rob as my supervisor. The continuous guidance and support that he has given me with his enthusiasm, optimism and wisdom have been the most important contributors to my Ph.D. I sincerely hope to keep working together for years to come.

In addition to my supervisors and thesis committee, my thanks go to the following people who contributed to this thesis in one way or another:

- Marcel, Umut, Sergio, Isabelle, Hugo and Luca for collaborating with me to explore all sorts of interesting things from affective computing to neural coding.
- Rob, Arno, Richard, Tessa, Chayada, Xuyan, Simon, Andrea, Jordy and other members of the Perception and Awareness Lab, and Marcel, Umut, Max, Sander, Linda, Pasi, Katja, Nadine, Silvan, Gabi and Erdi and other members of the Artificial Cognitive Systems Lab for the inspiring discussions, interesting presentations and helpful feedback.
- Marcel for being there for me (and Umut) for research and otherwise.
- Chayada for being my good friend and officemate, and helping me procrastinate.
- Jolanda and Vanessa for being very friendly and helpful whatever the situation, and Ronny for patiently counting Iris cheques and always having Presentation dongles for me.
- Anonymous test subjects for participating in my experiments.
- Inge and Luc for giving me very valuable teaching opportunities, Antonella and Lonneke for sharing these opportunities with me (mostly at 08:45 in the morning!), and Judith, Rathiga, Fleur and Lynn for being great students.
- DE Café personnel for supplying my daily dose of caffein, chocolate and panini.
- Joukje for the mentorship.

Finally, I would like to thank my dear partner, best friend and favourite colleague Umut for the fun, happiness and love, and the confidence, patience and support.

Anneme, babama, ablama, anneanneme, dedeme, babaanneme, büyük babama, teyzelerime, halama, Mine teyzeye, Kazım amcaya, Özlem'e, Caner'e, Yalın'a, Begüm'e, Güçlü'ye, Nehir'e, Günay teyzeye, Yurdanur amcaya, Erol abiye, Fikri amcaya, Ergin amcaya ve Cüneyt abiye bana olan destekleri ve sevgileri için; ayrıca, anneme ve babama benim için yaptıkları ve yazıya dökülmesi mümkün olmayan herşey için sonsuz teşekkürlerimi sunuyorum.

Curriculum vitae

Experience

- **Postdoctoral Researcher**
  Radboud University, Nijmegen, 2017 – present
- **Doctoral Researcher**
  Radboud University, Nijmegen, 2013 – 2017
- **Research Assistant**
  Radboud University, Nijmegen, 2012 - 2013

Education

- **Ph.D. Cognitive Neuroscience**
  Radboud University, Nijmegen, 2013 – 2017
  Supervisor: Dr. Rob van Lier; Thesis: Effects of complexity in perception: from construction to reconstruction
- **M.Sc. Cognitive Neuroscience**
  Radboud University, Nijmegen, 2011 – 2013
  Grade: 8.20/10.00 (thesis: 8.00/10.00); Honors: Cum laude; Supervisors: Prof. Ole Jensen and Dr. Mathilde Bonnefond; Thesis: Do alpha oscillations provide a mechanism for prioritizing salient unattended stimuli?
- **B.IT Artificial Intelligence**
  Multimedia University, Malacca, Malaysia, 2008 – 2011
  Grade: 3.84/4.00 (thesis: 4.00/4.00); Honors: First-class; Supervisor: Prof. Andrews Samraj; Thesis: Brain-computer interfacing: a study on the P300 speller paradigm

Awards and fellowships

- **3rd Place Award in Apparent Personality Analysis** (challenge award)
  ChaLearn Looking at People Challenge and Workshop at ECCV, Amsterdam, 2016
- **Academy Assistants Programme** (one-year research assistant fellowship)
  Royal Netherlands Academy of Arts and Sciences, Amsterdam, 2012
- **Huygens Scholarship Programme** (two-year master's fellowship)
  Netherlands Universities Foundation for International Cooperation, The Hague, 2011
- **Book Award** (bachelor's award)
  Multimedia University, Malacca, Malaysia, 2011

Publications and presentations

Preprint

- **Güçlütürk, Y.**, and van Lier, R. (2018). Decomposing complexity preferences for music. *PsyArXiv*.
- Ambrogioni, L., Güçlü, U., Berezutskaya, J., van den Borne, E., **Güçlütürk, Y.**, Hinne, M., Maris, E., and van Gerven, M. (2018). Forward Amortized Inference for Likelihood-Free Variational Marginalization. ***arXiv preprint arXiv:1805.11542 [stat.ML]***. https://arxiv.org/pdf/1805.11542.pdf (https://arxiv.org/pdf/1805.11542.pdf)
- Ambrogioni, L., Güçlü, U., , **Güçlütürk, Y.**, Hinne, M., van Gerven, M., and Maris, E. (2018). Wasserstein variational inference. ***arXiv preprint arXiv:1805.11284 [stat.ML]***. https://arxiv.org/pdf/1805.11284.pdf (https://arxiv.org/pdf/1805.11284.pdf)
- Jacques Junior, J., **Güçlütürk, Y.**, Pérez, M., Güçlü, U., Andujar, C., Baró, X., Escalante, H., Guyon, I., van Gerven, M., van Lier, R., Escalera, S. (2018). First impressions: a survey on computer vision-based apparent personality trait analysis. ***arXiv preprint arXiv:1804.08046 [cs.CV]***. https://arxiv.org/pdf/1804.08046.pdf (https://arxiv.org/pdf/1804.08046.pdf)
- Escalante, H., Kaya, H., Salah, A., Escalera, S., **Güçlütürk, Y.**, Güçlü, U., Baró, X., Guyon, I., Jacques Junior, J., Madadi, M., Ayache, S., Viegas, E., Gürpınar, F., Wicaksana, A., Liem, C., van Gerven, M., and van Lier, R. (2018). Explaining first impressions: modeling, recognizing, and explaining apparent personality from videos. ***arXiv preprint arXiv:1802.00745 [cs.CV]***. https://arxiv.org/pdf/1802.00745.pdf (https://arxiv.org/pdf/1802.00745.pdf)
- Seeliger, K., Güçlü, U., Ambrogioni, L., **Güçlütürk, Y.**, and van Gerven, M. (2017). Generative adversarial networks for reconstructing natural images from brain activity. ***bioRxiv***. https://doi.org/10.1101/226688 (https://doi.org/10.1101/226688)
- Güçlü, U., **Güçlütürk, Y.**, Madadi, M., Escalera, S., Baró, X., González, J., van Lier, R., and van Gerven, M. (2017). End-to-end semantic face segmentation with conditional random fields as convolutional, recurrent and adversarial networks. ***arXiv preprint arXiv:1703.03305 [cs.CV]***. https://arxiv.org/pdf/1703.03305.pdf (https://arxiv.org/pdf/1703.03305.pdf)

Journal

- **Güçlütürk, Y.**, Güçlü, U., van Gerven, M., and van Lier, R. (2018). Representations of naturalistic stimulus complexity in early and associative visual and auditory cortices. *Scientific Reports*, 8:3439. https://doi.org/10.1038/s41598-018-21636-y (https://doi.org/10.1038/s41598-018-21636-y)
- **Güçlütürk, Y.**, Güçlü, U., Baró, X., Escalante, H., Guyon, I., Escalera, S., van Gerven, M., and van Lier, R. (2017). Multimodal first impression analysis with deep residual networks. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2017.2751469 (https://doi.org/10.1109/TAFFC.2017.2751469)

- **Güçlütürk, Y.**, Jacobs, R., and van Lier, R. (2016). Liking versus complexity: decomposing the inverted U-curve. *Frontiers in Human Neuroscience*, 10:112. https://doi.org/10.3389/fnhum.2016.00112 (https://doi.org/10.3389/fnhum.2016.00112)

Conference

Paper

- Güçlü, U., Güçlütürk, Y., Ambrogioni, L., Maris, E., van Lier, R., and van Gerven, M. (2017). Algorithmic composition of polyphonic music with the wavecrf. In *Machine Learning for Creativity and Design - Neural Information Processing Systems*. https://nips2017creativity.github.io/doc/WaveCRF.pdf (https://nips2017creativity.github.io/doc/WaveCRF.pdf) ‡

- Ambrogioni, L., Berezutskaya, J., Güçlü, U., van den Borne, E., **Güçlütürk, Y.**, van Gerven, M., and Maris, E. (2017). Bayesian model ensembling using meta-trained recurrent neural networks. In *Workshop on Meta-Learning - Neural Information Processing Systems*. http://metalearning.ml/papers/metalearn17_ambrogioni.pdf (http://metalearning.ml/papers/metalearn17_ambrogioni.pdf) ‡

- **Güçlütürk, Y.**, Güçlü, U., Seeliger, K., Bosch, S., van Lier, R., and van Gerven, M. (2017). Reconstructing perceived faces from brain activations with deep adversarial neural decoding. In *Neural Information Processing Systems*. https://papers.nips.cc/paper/7012-reconstructing-perceived-faces-from-brain-activations-with-deep-adversarial-neural-decoding.pdf (https://papers.nips.cc/paper/7012-reconstructing-perceived-faces-from-brain-activations-with-deep-adversarial-neural-decoding.pdf) ‡ (**best poster award by Donders Institute**)

- **Güçlütürk, Y.**, Güçlü, U., Pérez, M., Escalante, H., Baró, X., Guyon, I., Andujar, C., Jacques Junior, J., Madadi, M., Escalera, S., van Gerven, M., and van Lier, R. (2017). Visualizing apparent personality analysis with deep residual networks. In *Action, Gesture, and Emotion Recognition Workshop and Competitions: Large Scale Multimodal Gesture Recognition and Real versus Fake Expressed Emotions - IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCVW.2017.367 (https://doi.org/10.1109/ICCVW.2017.367) ‡

- Escalante, H., Guyon, I., Escalera, S., Jacques Junior, J., Madadi, M., Baró, X., Ayache, S., Viegas, E., **Güçlütürk, Y.**, Güçlü, U., et al. (2017). Design of an explainable machine learning challenge for video interviews. In *IEEE International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2017.7966320 (https://doi.org/10.1109/IJCNN.2017.7966320) †

- **Güçlütürk, Y.**, Güçlü, U., van Gerven, M., and van Lier, R. (2016). Deep impression: audiovisual deep residual networks for multimodal apparent personality trait recognition. In *Challenge on Automatic Personality Analysis - European Conference on Computer Vision*. https://doi.org/10.1007/978-3-319-49409-8_28 (https://doi.org/10.1007/978-3-319-49409-8_28) †

- **Güçlütürk, Y.**, Güçlü, U., van Lier, R., and van Gerven, M. (2016). Convolutional sketch inversion. In *Computer Vision for Art Analysis - European Conference on Computer Vision*.

https://doi.org/10.1007/978-3-319-46604-0_56 (https://doi.org/10.1007/978-3-319-46604-0_56) †

Demo

- Escalera, S., Guyon, I., Chen, B., Quintana, M., Güçlü, U., **Güçlütürk, Y.**, Baró, X., van Lier, R., Andujar, C., van Gerven, M., Boser, B., and Wang, L. (2016). Biometric applications of CNNs: get a job at "Impending Technologies"!. In *Neural Information Processing Systems*. ‡

Abstract

- **Güçlütürk, Y.**, and Güçlü, U. (2018). Brain reading and writing with deep learning: future of neurotechnology. In *NVIDIA GPU Technology Conference Europe*. †
- Le, L., **Güçlütürk, Y.**, van Gerven, M., and van Wezel, R. (2018). Biological motion for bionic vision. In *Dutch Neuroscience Meeting*. ‡
- Güçlü, U., **Güçlütürk, Y.**, Ambrogioni, L., Maris, E., van Lier, R., and van Gerven, M. (2017). Ultrafast HRF and pRF estimation with deep learning. In *De Nederlandse Vereniging voor Psychonomie*. ‡
- **Güçlütürk, Y.** Perception, complexity and creativity: from old masters to learning machines. (2016). In *NIP Leergang Leven is Leren*, Nederlands Instituut van Psychologen. †
- **Güçlütürk, Y.**, Güçlü, U., Jacobs, R., van Gerven, M., and van Lier, R. (2016). Neural correlates of liking and perceived complexity of music. In *International Association of Empirical Aesthetics*. ‡
- **Güçlütürk, Y.**, Jacobs, R., and van Lier, R. (2015). Liking versus complexity: decomposing the inverted U-curve by accounting for individual differences. In *De Nederlandse Vereniging voor Psychonomie*. ‡
- **Güçlütürk, Y.**, Jacobs, R., and van Lier, R. (2015). Differential effects of task anticipation on liking of familiar surfaces. In *European Conference on Visual Perception*. ‡
- Jacobs, R., **Güçlütürk, Y.**, and van Lier, R. (2015). What you need is what you like: knowing target and distractor categories is sufficient for distractor devaluation. In *European Conference on Visual Perception*. ‡
- **Güçlütürk, Y.**, Jacobs, R., and van Lier, R. (2014). Perceptual complexity and liking of fractal-like statistical geometric patterns. In *European Conference on Visual Perception*. ‡
- Jacobs, R., **Güçlütürk, Y.**, and van Lier, R. (2014). Devaluation of distractors limited to evaluations after incorrect target localization. In *European Conference on Visual Perception*. ‡

Book

- Escalante, H., Escalera, S., Guyon, I., Baró, X., **Güçlütürk, Y.**, Güçlü, U., and van Gerven,

Marcel. (2018). ***Explainable and Interpretable Models in Computer Vision and Machine Learning***. Springer Science+Business Media.

*equal contribution; †oral presentation; ‡poster presentation

Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit: http://www.ru.nl/donders/graduate-school/phd/ (http://www.ru.nl/donders/graduate-school/phd/)